BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

BC

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/12, C07K 16/18, 14/47, A61K 38/17, G01N 33/68, 33/53, C12Q 1/68

A1

(11) International Publication Number: WO 95/02053

(43) International Publication Date: 19 January 1995 (19.01.95)

(21) International Application Number: PCT/US94/07297

(22) International Filing Date: 5 July 1994 (05.07.94)

(30) Priority Data:

08/088,483	6 July 1993 (06.07.93)	US
08/099,998	30 July 1993 (30.07.93)	US
08/113,971	30 August 1993 (30.08.93)	US
08/148,061	5 November 1993 (05.11.93)	US
08/222,626	4 April 1994 (04.04.94)	US

(71) Applicant: SCHERING CORPORATION [US/US]; 2000 Galloping Hill Road, Kenilworth, NJ 07033 (US).

(72) Inventors: ARAI, Naoko; 648 Georgia Avenue, Palo Alto, CA 94306 (US). MASUDA, Esteban, S.; 2250 Latham Street #29, Mountain View, CA 94040 (US). TOKUMITSU, Hiroshi; Apartment 8, 6108 S.W. 18th Drive, Portland, OR 97201 (US).

(74) Agents: BLASDALE, John, H., C. et al.; Schering-Plough Corporation, Patent Department K-6-1 1990, 2000 Galloping Hill Road, Kenilworth, NJ 07033-0530 (US).

(81) Designated States: AM, AU, BB, BG, BR, BY, CA, CN, CZ, FI, GE, HU, JP, KG, KR, KZ, LK, LT, LV, MD, MG, MN, NO, NZ, PL, RO, RU, SI, SK, TJ, TT, UA, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), ARIPO patent (KE, MW, SD).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: PURIFIED COMPONENTS OF MAMMALIAN TRANSCRIPTION REGULATION COMPLEXES, AND ANALOGS

(57) Abstract

Components of transcription regulation factors from a mammal, reagents related thereto including purified proteins, specific antibodies, and nucleic acids encoding said protein are provided. Methods of using said reagents and diagnostic kits are also provided.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgystan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo			SE	Sweden
CH	Switzerland	KR	Republic of Korea	SI	Slovenia
CI	Côte d'Ivoire	KZ	Kazakhstan	SK	Slovakia
CM	Cameroon	LI	Liechtenstein	SN	Senegal
CN	China	LK	Sri Lanka	TD	Chad
CS	Czechoslovakia	LU	Luxembourg	TG	Togo
CZ	Czech Republic	LV	Latvia	TJ	Tajikistan
DE	Germany	MC	Monaco	TT	Trinidad and Tobago
DK	Denmark	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	US	United States of America
FI	Finland	ML	Mali	UZ	Uzbekistan
FR	France	MN	Mongolia	VN	Viet Nam
GA	Gabon				

PURIFIED COMPONENTS OF MAMMALIAN TRANSCRIPTION REGULATION COMPLEXES, AND ANALOGS

FIELD OF THE INVENTION

The present invention provides compositions related to proteins which function in controlling development and differentiation of mammalian cells, e.g., cells of a mammalian immune system. In particular, it provides proteins and mimetics which regulate development, differentiation, and expression of various genes, including various cytokines.

BACKGROUND OF THE INVENTION

Rapid induction of cytokine production, including interleukin-2 (IL-2) and granulocyte-macrophage colony-stimulating factor (GM-CSF) by the stimulation of antigen-specific receptors of T lymphocytes, is an essential feature for acquired immune and inflammatory responses. Extensive studies have shown that antigen stimulation of the T-cell receptor triggers two types of signal transduction systems which result in induction of the lymphokine genes. One pathway is mediated by mobilization of intracellular Ca^{2+} promoted by phosphoinositol-turnover coupled to T-cell receptor/CD3 complex stimulation, and the other is the activation of protein kinase C. These two pathways, which are required for transcriptional activation of the lymphokine genes, can be mimicked by the Ca^{2+} ionophore A23187 and phorbol myristic acid (PMA). The immunosuppressive drugs cyclosporin A (CsA) and FK506 inhibit the stimulation elicited by A23187 and PMA.

The properties of one of the nuclear factors (here referred to as NF-CLE0γ) that binds to the Conserved Lymphokine gene Element 0 (CLE0) resembles that of the Nuclear Factor of Activated T cells (NF-AT), i.e. it exhibits inducibility by both PMA)and A23187 and sensitivity to CsA or cycloheximide.

The transcription regulation is mediated by quaternary protein complexes comprising a number of proteins. The AP-1 subcomponent of an NF-AT complex

is made up of two proteins, a representative of the jun family, and a representative of the fos family; see Jain et al. (1992) Nature 356:801-804. The lack of complete characterization of the other components of the NF-AT complex has led to the inability to understand the important components and interactions which regulate cytokine transcription and expression.

NF-AT is a lymphoid-specific transcription factor involved in regulation of the IL-2 gene and is considered to be an important regulator in early T cell activation. Recent studies have shown that NF-AT is a multimeric protein complex composed of the previously identified transcription factor AP-1 and a cytoplasmic component. See: Jain et al. (1992) Nature 356:801-804; Northrop et al. (1993) J. Biol. Chem. 268:2917-2923; and McCaffrey et al. (1993) J. Biol. Chem. 268:3747-3752. It has been reported that the appearance of the cytoplasmic component of NF-AT in the nucleus occurs in a Ca^{2+}-dependent manner and is blocked by CsA and FK-506 through inhibition of the Ca^{2+}/calmodulin-dependent protein phosphatase, calcineurin. However, the component of NF-AT which is associated with AP-1 has not been well characterized biochemically.

Isolation of these components is difficult because the amounts are very low within cells. Biochemical characterization has been blocked by the inability to produce sufficient amounts of purified proteins to perform *in vitro* studies. Thus, the mechanisms of transcriptional regulation of cytokines have been poorly understood. As cytokine and immune regulation are closely related and are basic features of immune physiology, understanding the control mechanisms will reap great benefits for medically relevant abnormalities in immune responsiveness.

SUMMARY OF THE INVENTION

The present invention is based, in part, upon the isolation of various components of a complex of proteins which interact to function as a transcription regulatory entity. The purified component, designated NF-AT120, interacts with AP-1 proteins (comprising a member of each of the jun and fos protein families) to form a multiprotein complex which binds to an NF-AT recognition sequence. The invention identifies a common factor that regulates transcriptional activation of the GM-CSF and IL-2 genes upon T cell activation. In fact, there exist at least three closely related subfamilies of NF-AT proteins, designated herein as type C, type P, and type X subfamilies. Each subfamily consists of closely related variants.

The invention embraces purified natural forms as well as analogues and homologues, e.g., mutations (muteins) of the natural sequence, fusion proteins, chemical mimetics, antibodies, and other structural or functional analogues. It is also directed to isolated genes encoding proteins of the invention. Various uses of these different protein or nucleic acid compositions are also provided.

The present invention provides a substantially pure component of an NF-AT protein complex, or peptide thereof, or a fusion protein comprising NF-AT120 protein sequence; an antibody specific for binding to an NF-AT120 protein; and a nucleic acid encoding an NF-AT120 protein or fragment thereof.

In embodiments encompassing a substantially pure NF-AT120 protein or peptide thereof, the protein or peptide can be from a warm blooded animal selected from the group of birds and mammals, including a mouse, rat, or human; can comprise at least one polypeptide segment of SEQ ID NO: 1 through 5, 35, 37, 39 and 41; can exhibit a post-translational modification pattern distinct from natural NF-AT120 protein; can exhibit at least one of the features disclosed in Table 1; or can induce transcription of a cytokine. (All SEQ IDs are given together immediately before the Claims.) A further embodiment is a composition comprising such a protein and a pharmaceutically acceptable carrier.

In antibody embodiments, the antigen can be a mammalian protein, e.g., from a mouse, rat, or human; the antibody can be raised against a peptide sequence of SEQ ID NO: 1 through 5, 35, 37, 39 and 41; the antibody can be a monoclonal antibody; or the antibody can be labeled.

In nucleic acid embodiments, the nucleic acid can comprise a sequence of SEQ ID NO: 6 through 24, 34, 36, 38 and 40.

The invention also embraces a kit comprising a substantially pure NF-AT120 protein or fragment, e.g., as a positive control; an antibody which specifically binds an NF-AT120 protein; or a nucleic acid encoding an NF-AT120 protein or peptide.

The availability of these reagents also provides methods of modulating physiology or development of a cell comprising contacting said cell with an NF-AT120 protein or with an analogue or homologue thereof, e.g., by introducing said NF-AT120 protein or analogue or homologue thereof into a cell. For example, an inhibitor might be an antibody against a mammalian NF-AT120 protein or the cell may be a hematopoietic cell, including a lymphoid cell.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows a mobility shift assay of affinity-purified NF-AT. Nuclear extracts of PMA/A23187-stimulated Jurkat cells (NE) were assayed in the presence of 250 ng/µl poly-dIdC (lanes 1 and 5). Mobility shift assays of affinity-purified NF-AT (30 ng) were carried out without poly-dIdC (lanes 2-4, and 6-11) using ^{32}P-labeled NF-AT and CLE0 oligonucleotide as probes. The probes are listed at the top with the competitors identified below each probe.

Figure 2 shows a purification of the 120 kDa component of NF-AT (NF-AT120) by Mono Q chromatography. Affinity-purified NF-AT was applied in the presence of 6 M urea and eluted by KCl gradient (0.05 - 0.8 M) from a Mono Q column. Mobility shift assays were carried out in a solution containing 1 µl of each fraction, 100 ng/µl polydIdC in the absence (Figure 2A) or presence (Figure 2B) of 10 ng of affinity-purified Jurkat AP-1. The right lane in Figure 2B shows NF-AT binding of 10 ng of affinity-purified Jurkat AP-1. Figure 2C shows analysis by SDS-7.5% PAGE (polyacrylamide gel electrophoresis) of each fraction from Mono Q chromatography. The left lane (M) indicates molecular weight standards (kDa). Fraction numbers are listed below each figure.

Figure 3 shows that recombinant cJun/cFos heterodimer reconstitutes NF-AT DNA-binding with the NF-AT120 protein. Mobility shift assays were carried out using NF-AT DNA probe in a solution containing 10 ng of affinity-purified Jurkat AP-1 (lane 2) or various combinations of recombinant cJun (0.2 µM) and cFos (0.2 µM) in either the presence (+) or absence (–) of the Mono Q-purified 120 kDa protein (1 µl). The proteins are listed on top.

Figure 4 shows that NF-AT120 reconstitutes the DNA-binding activity to NF-AT and CLE0 elements with AP-1. Figure 4A shows analysis by SDS-7.5 % PAGE of purified NF-AT120 (lane 1). The left lane (M) indicates molecular weight standards (kDa). Figure 4B shows mobility shift assays carried out using NF-AT DNA probe and CLE0 DNA probe. The reaction solutions contain the 120 kDa protein from Mono Q fraction ((Q), lanes 2, 3, 7, 8,) or renatured NF-AT120 protein from gel slice ((G), lanes 4, 5, 9, 10) in either the presence (+AP-1) or absence of 10 ng of affinity-purified Jurkat AP-1. NF-AT DNA binding of affinity-purified NF-AT (30 ng) is indicated in lane 1.

Figure 5 is a diagrammatic representation showing how proteins of the present invention are currently believed to bind to dsDNA. However, the validity of the claims of the present invention is not dependent upon the correctness of this belief. In this Figure, PKC represents protein kinase C, CHX represents cyclohexamide, A is AP-1 as defined herein, CsA is cyclosporin A, and B

(NF-ATc) represents the C family of NF-AT, which is further defined in the Sequence Listing.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

GENERAL

The present invention provides the amino acid sequences of and DNA sequences encoding various mammalian proteins which exhibit properties of regulating transcription of immunologically relevant proteins, e.g., cytokines. These proteins are designated components of a Nuclear Factor of Activated T cells (NF-AT) because they were initially characterized as proteins which associate in a complex and regulate transcription of various cytokines, e.g., IL-2 and GM-CSF. The proteins are one component, referred to herein as NF-AT120, and exhibit features characteristic of a transcriptional regulatory factor component, e.g., specific association with particular genetic regulatory sequences when associated with other components, i.e., AP-1, of the complex. The best characterized embodiment was initially described in humans, and various subclasses have been identified herein. Similar sequences for proteins in other mammalian species, e.g., monkeys, rats, and mice, should also be available. The descriptions below are directed, for exemplary purposes, to human NF-AT120 proteins, but are likewise applicable to related embodiments from other species.

Purified NF-AT120 proteins

The present invention provides substantially purified NF-AT120 proteins, derived from either natural sources or recombinant sources. The proteins exhibit properties as described, both physicochemical and biological. See Table 1:

Table 1: Physical properties of human NF-AT120 protein.

SDS-Polyacrylamide gel electrophoresis: reduced migration approximately 120 Kd.

Precipitation of the NF-AT complex with ammonium sulfate (at 4°C.): found in 20-40% saturated $(NH_4)_2SO_4$ pellet.

Affinity binding of NF-AT complex to a CLE0 DNA sequence or NF-AT DNA binding sequence, but competed by oligonucleotides containing NF-AT DNA binding and AP-1 DNA binding sites.

Reconstitution with AP-1 (recombinant cJun/cFos heterodimer) provides NF-AT complex.

Anion Exchange Chromatography (KCl gradient in 6 M urea, 50 mM KCl, 20 mM HEPES (pH 7.9), 1 mM EGTA, 1 mM EDTA) on Mono Q column): activity eluted between 200 and 300 mM KCl.

Gel Filtration (SUPEROSE 6 column, Pharmacia HR16/50): the NF-AT complex ran with an apparent molecular weight of 440 kD, in 0.1 M KCl, 20 mM HEPES (pH 7.9), 1 mM EDTA, 1 mM EGTA, 1 mM DTT, 0.1 mM PMSF, 10 µg/ml leupeptin, 5% glycerol.

NF-AT DNA binding affinity chromatography in 0.2% NP-40, washed with 1 M urea, elutes in 0.3 M KCl without urea.

Lys C protease fragments of a purified human NF-AT120 protein amino acid sequence are presented as SEQ ID NO: 1 through 5, reading from the amino to the carboxy end. Other peptide sequences are provided by sequences from cloned nucleic acids encoding the proteins; see SEQ ID NO: 35, 37, 39, and 41. These amino acid sequences are important in providing partial sequence information in the protein, to enable the protein to be distinguished from other proteins. Moreover, the peptide sequences allow preparation of peptides to generate antibodies that recognize such segments, and/or allow preparation of oligonucleotide probes, both of which are strategies for isolation, e.g., cloning, of genes encoding such sequences. In addition, another peptide-sequencing experiment showed two peptides, whose overlapped sequence reads, by amino acid cycle:

1:V,P; 2:A,Y; 3:S,N; 4:P,P; 5:P,L; 6:A,S; 7:G,S; 8:P,L; 9:A,S; 10:Y,G; 11:P,E; 12:P,D; 13:D,P; 14:G,V; 15:R,L; 16:D,F; 17:G,Y; 18:E,P; 19:P,L; 20:D,K; 21:R.

As used herein, the term "human NF-AT120 protein" shall encompass, when used in a protein context, a protein containing at least some of the amino acid sequences of SEQ ID NO: 1 through 5, 35, 37, 39, or 41, or a significant fragment of such a protein, or protein sequences encoded by isolated NF-AT120 genes or transcripts. SEQ ID NO: 1 corresponds to residues 389-395 of SEQ ID NO: 36; SEQ ID NO: 3 corresponds to residues 279-287; residues 2-15 of SEQ ID NO: 4 correspond to residues 415-428; and residues 1-5 of SEQ ID NO: 5 correspond to residues 677-681. SEQ ID NO: 5 may be encoded by an alternative splicing variant of a message.

The term also refers to a human-derived polypeptide which exhibits similar biological function or interacts with NF-AT120 protein-specific binding components, including AP-1 proteins. The specific binding components, which include antibodies, typically bind to an NF-AT120 protein with high affinity, e.g., at better than about 100 nM, usually at better than about 30 nM or about 10 nM, and more preferably at better than about 3 nM. Under certain circumstances, the binding may require additional physiologically relevant cofactors or components. Homologous proteins would be found in mammalian species other than human, e.g., in other primates, in rats, and in mice. Non-mammalian species should also possess structurally or functionally related genes and proteins.

The term "polypeptide" as used herein includes a significant fragment or segment, and encompasses a stretch of amino acid residues of at least about 8 amino acids, generally at least 12 or 16 amino acids, preferably at least 20 or 24 amino acids, and, in particularly preferred embodiments, at least 28 or even 30 or more amino acids.

The term "binding agent" or "binding composition" refers to molecules that bind with specificity to NF-AT120. One embodiment includes antibodies that bind specifically to NF-AT120. A second embodiment includes proteins, like protein complex AP-1, that complex specifically with NF-AT120 protein or specifically associate with it, as in a natural physiologically relevant protein-protein interaction. The binding agent or binding composition may be a polymer or a chemical reagent. A functional analog thereto may be a protein with structural modifications, or may be a wholly unrelated molecule, e.g., one which has a molecular shape which interacts with the appropriate binding determinants, preferably exhibiting similar biological properties.

Solubility of a polypeptide or fragment depends upon the environment and the polypeptide. Various soluble fragments of the NF-AT120 are also included. Many variables affect polypeptide solubility, including temperature,

electrolyte environment, size and molecular characteristics of the polypeptide, and nature of the solvent. Typically, the temperature at which the polypeptide is used ranges from about 4°C to about 65°C. Usually the temperature at use is higher than about 18°C and more usually higher than about 22°C. For diagnostic purposes, the temperature will usually be about room temperature or higher, but less than the denaturation temperature of components in the assay. For therapeutic purposes, the temperature will usually be body temperature, typically about 37°C for humans, though under certain situations the temperature may be raised or lowered *in situ* or *in vitro*.

The electrolytes will usually approximate *in situ* physiological conditions, but may be modified to higher or lower ionic strength where advantageous. The actual ions may be modified, e.g., to conform to standard buffers used in physiological or analytical contexts.

The size and structure of the polypeptide should generally be in a substantially stable state, and usually not in a denatured state. The polypeptide may be associated with other polypeptides in a quaternary structure, e.g., to form an NF-AT complex, or to confer solubility, or associated with lipids or detergents in a manner which approximates natural lipid bilayer interactions. Physiologically, the NF-AT120 typically associates with AP-1 proteins to confer many of its natural biological functions.

The solvent will usually be a biologically compatible buffer, of a type used for preservation of biological activities, and will usually approximate a physiological solvent. Usually the solvent will have a pH near neutrality, typically between about 5 and 10, and preferably about 7.5. On some occasions, a detergent will be added, typically a mild non-denaturing one, e.g., CHS or CHAPS, in a concentration low enough to avoid significant disruption of structural or physiological properties of the protein.

Solubility is reflected by sedimentation measured in Svedberg units, which are a measure of the sedimentation velocity of a molecule under particular conditions. The determination of the sedimentation velocity was classically performed in an analytical ultracentrifuge, but is typically now performed in a standard ultracentrifuge; see: Freifelder (1982) Physical Biochemistry (2d ed.), W.H. Freeman; and Cantor and Schimmel (1980) Biophysical Chemistry, parts 1-3, W.H. Freeman & Co., San Francisco; each of which is hereby incorporated herein by reference. As a crude determination, a sample containing a putatively soluble polypeptide is spun in a standard full sized ultracentrifuge at about 50K rpm for about 10 minutes, and soluble molecules will remain in the supernatant.

A soluble particle or polypeptide will typically be less than about 30S, usually less than about 15S or even 10S, preferably less than about 6S, and, in particular preferred embodiments, less than about 4S or even 3S.

The NF-AT120 protein exhibits important structural and physical characteristics which confer an important biological function. The first is an ability to associate with other components, e.g., AP-1 proteins, to form a complex having affinity for specific DNA sequences. This specificity of binding provides a biological function in regulation of both IL-2 and GM-CSF cytokine expression. These two features have been utilized to isolate NF-AT120.

Physical Variants

This invention also encompasses proteins or peptides having substantial amino acid sequence homology with the amino acid sequences of the NF-AT120 protein. The variants include species or allelic variants.

Amino acid sequence homology, or sequence identity, is determined by optimizing residue matches, if necessary introducing gaps as required. This changes when conservative substitutions are considered to be matches. Conservative substitutions typically include substitutions within the following groups: glycine, alanine; valine, isoleucine, leucine; aspartic acid, glutamic acid; asparagine, glutamine; serine, threonine; lysine, arginine; and phenylalanine, tyrosine. Homologous amino acid sequences are typically intended to include natural allelic and interspecies variations in each respective protein sequence.

Typical homologous proteins or peptides will have from 25-100% homology (if gaps can be introduced) to 50-100% homology (if conservative substitutions are included) with the amino acid sequence of the NF-AT120 protein. Homology measures will be at least about 35%, generally at least 45%, often at least 55%, typically at least 65%, usually at least 75%, preferably at least 80%, and in particularly preferred embodiments, at least 85% or more. See also: Needleham et al. (1970) J. Mol. Biol., 48:443-453; Sankoff et al. (1983) Chapter One in Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison, Addison-Wesley, Reading, MA; and software packages from IntelliGenetics, Mountain View, CA, and from the University of Wisconsin Genetics Computer Group, Madison, WI.

The isolated NF-AT120 protein DNA can be readily modified by nucleotide substitutions, nucleotide deletions, nucleotide insertions, and inversions of nucleotide stretches. These modifications result in novel DNA

sequences which encode these antigens, their derivatives, or proteins having similar physiological, immunogenic, or antigenic activity. These modified sequences can be used to produce mutant antigens or to enhance expression. Enhanced expression may involve gene amplification, increased transcription, increased translation, and other mechanisms. Such mutant NF-AT120 protein derivatives include predetermined or site-specific mutations of the respective protein or its fragments. "Mutant NF-AT120 protein" encompasses a polypeptide otherwise falling within the homology definition of the human NF-AT120 protein as set forth above, but having an amino acid sequence which differs from that of NF-AT120 protein as found in nature, whether by way of deletion, substitution, or insertion. In particular, "site specific mutant NF-AT120 protein" generally includes proteins having significant homology with a protein having sequences of SEQ ID NO: 1 through 5, 35, 37, 39 or 41, and sharing various biological activities, e.g., antigenic or immunogenic, with those sequences, and preferred embodiments thereof contain most of the disclosed sequences. Similar concepts apply to different NF-AT120 proteins, particularly those found in various warm blooded animals, e.g., mammals and birds. As stated before, it is emphasized that descriptions are generally meant to encompass all NF-AT120 proteins, not limited to the embodiment specifically discussed.

Although site-specific mutation sites are predetermined, mutants need not be site specific. NF-AT120 protein mutagenesis can be conducted by making amino acid insertions or deletions. Substitutions, deletions, insertions, or any combinations may be generated to arrive at a final construct. Insertions include amino- or carboxy-terminal fusions. Random mutagenesis can be conducted at a target codon and the expressed mutants can then be screened for the desired activity. Methods for making substitution mutations at predetermined sites in DNA having a known sequence are well known in the art, e.g., by M13 primer mutagenesis or polymerase chain reaction (PCR) techniques; see also Sambrook et al. (1989) and Ausubel et al. (1987 and Supplements).

The mutations in the DNA normally should not place coding sequences out of reading frames and preferably will not create complementary regions that could hybridize to produce secondary mRNA structure such as loops or hairpins.

The present invention also provides recombinant proteins, e.g., heterologous fusion proteins using segments from these proteins. A heterologous fusion protein is a fusion of proteins or segments which naturally are not normally fused in the same manner. Thus, the fusion product of an immunoglobulin with an NF-AT120 protein polypeptide is a continuous protein

molecule having sequences fused in a typical peptide linkage, typically made as a single translation product and exhibiting properties derived from each source peptide. A similar concept applies to heterologous nucleic acid sequences.

In addition, new constructs may be made by combining similar functional domains from other proteins. For example, DNA-binding or other segments may be "swapped" between different new fusion polypeptides or fragments; see, e.g.: Cunningham et al. (1989) Science, 243:1330-1336; and O'Dowd et al. (1988) J. Biol. Chem., 263:15985-15992, each of which is incorporated herein by reference. Thus, new chimeric polypeptides exhibiting new combinations of specificities will result from the functional linkage of DNA-binding specificities or protein-binding specificities and other functional domains.

The phosphoramidite method described by Beaucage and Carruthers in (1981) Tetra. Letts. 22:1859-1862, will produce suitable synthetic DNA fragments. A double-stranded fragment will often be obtained either by synthesizing the complementary strand and annealing the strand together under appropriate conditions or by adding the complementary strand using DNA polymerase with an appropriate primer sequence, e.g., by PCR techniques.

Functional Variants

The blocking of physiological response to NF-AT120 proteins may result from the inhibition of binding of the protein to either DNA or other components of the NF-AT complex, likely through competitive inhibition. Thus, *in vitro* assays of the present invention will often use isolated protein, membranes from cells expressing a recombinant membrane associated NF-AT120 protein, soluble fragments comprising DNA or AP-1 protein binding segments of these proteins, or fragments attached to solid phase substrates. These assays will also allow for the diagnostic determination of the effects of either DNA- or protein-binding segment mutations and modifications, or compounds which can disrupt or facilitate such interactions.

This invention also contemplates the use of competitive drug screening assays, e.g., where neutralizing antibodies to NF-AT120 or AP-1 protein fragments compete with a test compound for binding to the protein. In this manner, the antibodies can be used to detect the presence of any polypeptide which shares one or more antigenic binding sites of the NF-AT120 and can also be used to occupy binding sites on the protein that might otherwise interact with another component of an NF-AT complex.

Additionally, neutralizing antibodies against the NF-AT120 protein and soluble fragments of the protein which contain a high affinity binding site to specific targets can be used to inhibit transcriptional function in tissues, e.g., tissues experiencing abnormal physiology.

"Derivatives" of the NF-AT120 protein antigens include amino acid sequence mutants, glycosylation variants, and covalent or aggregate conjugates with other chemical moieties. Covalent derivatives can be prepared by linkage of functionalities to groups which are found in the NF-AT120 protein amino acid side chains or at the N- or C-termini, by means which are well known in the art. These derivatives can include, without limitation, aliphatic esters or amides of the carboxyl terminus, or of residues containing carboxyl side chains, O-acyl derivatives of hydroxyl group-containing residues, and N-acyl derivatives of the amino terminal amino acid or amino-group containing residues, e.g., lysine or arginine. Acyl groups are preferably selected from the group of alkanoyl moieties including C2 to C18 normal alkanoyl. Covalent attachment to carrier proteins may be important when immunogenic moieties are haptens.

In particular, glycosylation alterations are included, e.g., made by modifying the glycosylation patterns of a polypeptide during its synthesis and processing, or in further processing steps. Particularly preferred means for accomplishing this are by exposing the polypeptide to glycosylating enzymes derived from cells which normally provide such processing, e.g., mammalian glycosylation enzymes. Deglycosylation may be effected by appropriate enzymes. Also embraced are versions of the same primary amino acid sequence which have other minor modifications, including phosphorylated amino acid residues, e.g., phosphotyrosine, phosphoserine, or phosphothreonine.

A major group of derivatives comprises covalent conjugates of the NF-AT120 protein or fragments thereof with other proteins or polypeptides. These derivatives can be synthesized in recombinant culture such as N- or C-terminal fusions or by the use of agents known in the art for their usefulness in cross-linking proteins through reactive side groups. Preferred protein derivatization sites with cross-linking agents are at free amino groups, carbohydrate moieties, and cysteine residues.

Fusion polypeptides between the NF-AT120 proteins and other homologous or heterologous proteins are also provided. Homologous polypeptides may be fusions between different nuclear proteins, resulting in, e.g., a hybrid protein exhibiting DNA binding specificity or nuclear factor binding

specificity. Likewise, heterologous fusions may be constructed which would exhibit a combination of properties or activities of the derivative proteins. Typical examples are fusions of a reporter polypeptide, e.g., luciferase, with a segment or domain of a protein, e.g., a DNA-binding segment, so that the presence or location of the fusion protein may be easily determined; see, e.g., Dull et al., U.S. Patent No. 4,859,609. Other gene fusion partners include bacterial β-galactosidase, trpE, Protein A, β-lactamase, alpha amylase, alcohol dehydrogenase, and yeast alpha mating factor; see, e.g., Godowski et al. (1988) Science 241:812-816.

Some polypeptides may also have amino acid residues which have been chemically modified by phosphorylation, sulfonation, biotinylation, or the addition or removal of other moieties, particularly those which have molecular shapes similar to phosphate groups. In some embodiments, the modifications will be useful labeling reagents, or serve as purification targets, e.g., affinity ligands, or fusions allowing for simple purification and processing.

Fusion proteins will typically be made either by recombinant nucleic acid methods or by synthetic polypeptide methods. Techniques for nucleic acid manipulation and expression are described generally in, for example, Sambrook et al. (1989) Molecular Cloning: A Laboratory Manual (2d ed.), Vols. 1-3, Cold Spring Harbor Laboratory. Techniques for synthesis of polypeptides are described in, for example: Merrifield (1963) J. Amer. Chem. Soc., 85:2149-2156; Merrifield (1986) Science 232:341-347; and Atherton et al. (1989) Solid Phase Peptide Synthesis: A Practical Approach, IRL Press, Oxford.

This invention also contemplates the use of derivatives of the NF-AT120 proteins other than variations in amino acid sequence or glycosylation. Such derivatives may involve covalent or aggregative association with chemical moieties. These derivatives generally fall into the three classes: (1) salts, (2) side-chain covalent modifications and terminal residue covalent modifications, and (3) adsorption complexes, for example with cell membranes. Such covalent or aggregative derivatives are useful as immunogens, as reagents in immunoassays, or in purification methods such as affinity purification of binding partners. For example, an NF-AT120 protein antigen can be immobilized by covalent bonding to a solid support such as cyanogenbromide-activated SEPHAROSE, by methods which are well known in the art, or adsorbed onto polyolefin surfaces, with or without glutaraldehyde cross-linking, for use in the assay or purification of anti-NF-AT120 protein antibodies or its binding partners. The NF-AT120 proteins can also be labeled with a detectable

group, for example radioiodinated by the chloramine T procedure, covalently bound to rare earth chelates, or conjugated to another fluorescent moiety for use in diagnostic assays. Purification of NF-AT120 protein may be effected by immobilized antibodies or binding partners.

A solubilized NF-AT120 protein or fragment of this invention can be used as an immunogen for the production of antisera or antibodies specific for the protein or any fragments thereof. The purified proteins can be used to screen monoclonal antibodies or antigen-binding fragments prepared by immunization with various forms of impure preparations containing the protein. In particular, the term "antibodies" also encompasses antigen-binding fragments of natural antibodies. The purified NF-AT120 proteins can also be used as a reagent to detect any antibodies generated in response to the presence of elevated levels of the protein, protein complex, or cell fragments containing the protein, each of which may be diagnostic of an abnormal or specific physiological or disease condition, e.g., autoimmunity. Additionally, protein fragments may also serve as immunogens to produce the antibodies of the present invention, as described immediately below. For example, this invention contemplates antibodies raised against amino acid sequences shown in Table 1, or fragments of proteins containing them.

The present invention contemplates the isolation of additional closely related species variants. Southern, Northern, and Western blot analysis should establish that similar genetic entities exist in other mammals. It is likely that the NF-AT120 proteins are widespread in various orders and species, e.g., rodents, lagomorphs, carnivores, artiodactyla, perissodactyla, and primates.

The invention also provides means to isolate a group of related antigens displaying both distinctness and similarities in structure, expression, and function. Elucidation of many of the physiological effects and mechanism of action of the proteins will be greatly accelerated by the isolation and characterization of distinct species variants of the proteins. In particular, the present invention provides useful probes for identifying additional homologous genetic entities in different species.

The isolated genes will allow transformation of cells lacking expression of a corresponding NF-AT120 protein, e.g., either species types or cells which lack corresponding proteins and exhibit negative background activity. Expression of transformed genes will allow isolation of antigenically pure cell lines, with defined or single species variants. This approach will allow for more sensitive detection and discrimination of the physiological effects of proteins regulating

nuclear transcription. Subcellular fragments, e.g., cytoplasts or membrane fragments, can be isolated and used in appropriate assays.

Dissection of the critical structural elements which effect the various differentiation functions provided by this factor is possible using standard techniques of modern molecular biology, particularly in comparing members of the related class. See, e.g.: the homolog-scanning mutagenesis technique described by Cunningham et al. (1989) Science, 243:1339-1336; and approaches used by O'Dowd et al. (1988) J. Biol. Chem., 263:15985-15992; and by Lechleiter et al. (1990) EMBO J., 9:4381-4390; each of which is incorporated herein by reference.

In particular, DNA-binding segments or AP-1 binding segments can be substituted between species variants to determine what structural features are important in both DNA binding affinity and specificity, as well as transcriptional activation. An array of different NF-AT120 variants will be used to screen for factors exhibiting combined properties of interaction with different DNA or AP-1 species variants.

Various functions would probably involve segments of the factor which are normally accessible to DNA or protein binding. Dissection of what domains of the protein are involved in DNA binding or AP-1 protein interaction can be performed, along with elucidation of other NF-AT components. The specific segments of interaction of NF-AT120 protein with other components may be identified by mutagenesis or direct biochemical means, e.g., cross-linking or affinity methods. Structural analysis by crystallographic or other physical methods will also be applicable. Further investigation of the mechanism of activation or suppression of transcription will include study of associated components which may be isolatable by affinity methods or by genetic means, e.g., complementation analysis of mutants.

Further study of the expression and control of NF-AT120 protein will be pursued. The controlling elements associated with the factors may exhibit differential developmental, tissue-specific, or other expression patterns. Upstream or downstream genetic regions, e.g., control elements, are of interest. Structural studies of the factors will lead to design of new factors, particularly analogues exhibiting constitutive suppression of activation of transcription. This can be combined with previously described screening methods to isolate proteins exhibiting desired spectra of activities.

Expression in other cell types will often result in different glycosylation in a particular factor. Various species variants may exhibit distinct functions based

upon structural differences other than amino acid sequence. Differential modifications may be responsible for differential function, and elucidation of the effects are now made possible.

Thus, the present invention provides important reagents related to a physiological NF-AT120 interaction with AP-1 and/or regulatory DNA segments. Although the foregoing description has focused primarily upon a human NF-AT120 protein, those of skill in the art will immediately recognize that the invention encompasses other species variants, e.g., rat and other mammalian species or allelic variants, as well as other variants thereof.

Antibodies

Antibodies can be raised to the various NF-AT120 proteins, including species or allelic variants, and fragments thereof, both in their naturally occurring forms and in their recombinant forms. Additionally, antibodies can be raised to NF-AT120 proteins either in their active forms or in their inactive forms. Anti-idiotypic antibodies are also contemplated.

Antibodies, including binding fragments, e.g., Fab, Fab^2, Fv, etc., and single-chain versions, against predetermined fragments of the proteins can be raised by immunization of animals with conjugates of the fragments with immunogenic proteins. Monoclonal antibodies are isolated from cells secreting the desired antibody. These antibodies can be screened for binding to normal or mutant NF-AT120 proteins, or screened for suppressive or helper activity. These monoclonal antibodies will usually bind with a K_D of at most about 1 mM but more usually with stronger binding, e.g., with a K_D of at most about 300 μM, typically at most about 10 μM, more typically at most about 30 μM, preferably at most about 10 μM, and more preferably at most about 3 μM or better.

The antibodies, including antigen-binding fragments, e.g., Fab, Fab^2, Fv, etc., of this invention can have significant diagnostic or therapeutic value. They can be potent inhibitors that bind to the protein or complex and inhibit a biological response. They also can be useful as non-neutralizing antibodies and can be coupled to toxins or radionuclides so that, when the antibody binds to the protein, which can also be expressed on a cell surface or secreted into the proximal environment, the cell or nearby cells are killed. Further, these antibodies can be conjugated to drugs or other therapeutic agents, either directly or indirectly by means of a linker, and may effect drug targeting or cell labeling.

The antibodies of this invention can also be useful in diagnostic applications. As capture or non-neutralizing antibodies, they can be screened

for ability to bind to the proteins without inhibiting biological function. As neutralizing antibodies, they can be useful in competitive binding assays. They will also be useful in detecting or quantifying NF-AT120 protein or its complexes. See, e.g., Chan (ed.) (1987) Immunoassay: A Practical Guide, Academic Press, Orlando, FL; Price and Newman (eds.) (1991) Principles and Practice of Immunoassay, Stockton Press, N.Y.; and Ngo (ed.) (1988) Nonisotopic Immunoassay, Plenum Press, N.Y.

Protein fragments may be joined to other materials, particularly polypeptides, as fused or covalently joined polypeptides to be used as immunogens. A protein or its fragments may be fused or covalently linked to a variety of immunogens, such as keyhole limpet hemocyanin, bovine serum albumin, tetanus toxoid, etc. See: Microbiology, Hoeber Medical Division, Harper and Row, 1969; Landsteiner (1962) Specificity of Serological Reactions, Dover Publications, New York, and Williams et al. (1967) Methods in Immunology and Immunochemistry, Vol. 1, Academic Press, New York, for descriptions of methods of preparing polyclonal antisera. A typical method involves hyperimmunization of an animal with an antigen. The blood of the animal is then collected shortly after the repeated immunizations and the gamma globulin is isolated.

It is sometimes desirable to prepare monoclonal antibodies from various mammalian hosts, such as mice, rodents, primates, humans, etc. Description of techniques for preparing such monoclonal antibodies may be found in, e.g., Stites et al. (eds.) Basic and Clinical Immunology (4th ed.), Lange Medical Publications, Los Altos, CA, and references cited therein; Harlow and Lane (1988) Antibodies: A Laboratory Manual, CSH Press; Goding (1986) Monoclonal Antibodies: Principles and Practice (2d ed.) Academic Press, New York; and particularly in Kohler and Milstein (1975) in Nature 256:495-497, which discusses one method of generating monoclonal antibodies. Summarized briefly, this method involves injecting an animal with an immunogen. The animal is then sacrificed and cells taken from its spleen, which are then fused with myeloma cells. The result is a hybrid cell or "hybridoma" that is capable of reproducing *in vitro*. The population of hybridomas is then screened to isolate individual clones, each of which secretes a single antibody species to the immunogen. In this manner, the individual antibody species obtained are the products of immortalized and cloned single B cells from the immune animal generated in response to a specific site recognized on the immunogenic substance.

Other suitable techniques involve *in vitro* exposure of lymphocytes to the antigenic polypeptides or alternatively to selection of libraries of antibodies in phage or similar vectors. See: Huse et al. (1989) "Generation of a Large Combinatorial Library of the Immunoglobulin Repertoire in Phage Lambda," Science 246:1275-1281; and Ward et al. (1989) Nature 341:544-546. The polypeptides and antibodies of the present invention may be used with or without modification, including chimeric or humanized antibodies. Frequently, the polypeptides and antibodies will be labeled by joining, either covalently or non-covalently, a substance which provides for a detectable signal. A wide variety of labels and conjugation techniques are known and are reported extensively in both the scientific and patent literature. Suitable labels include radionuclides, enzymes, substrates, cofactors, inhibitors, fluorescent moieties, chemiluminescent moieties, magnetic particles, and the like. Patents teaching the use of such labels include: U.S. Patents Nos. 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149; and 4,366,241. Also, recombinant immunoglobulins may be produced; see Cabilly, U.S. Patent No. 4,816,567.

The antibodies of this invention can also be used for affinity chromatography in isolating proteins. Columns can be prepared where the antibodies are linked to a solid support, e.g., particles, such as agarose, SEPHADEX, or the like, where a cell lysate may be passed through the column, and the column is washed and then eluted with increasing concentrations of a mild denaturant, whereby the purified NF-AT120 protein or NF-AT will be released.

The antibodies may also be used to screen expression libraries for particular expression products. Usually the antibodies used in such a procedure will be labeled with a moiety allowing easy detection of presence of antigen by antibody binding.

Antibodies raised against an NF-AT120 protein will also be useful to raise anti-idiotypic antibodies. These will be useful in detecting or diagnosing various immunological conditions related to expression of the respective antigens.

Nucleic Acids

The described peptide sequences and the related reagents are useful in isolating a DNA clone encoding NF-AT120 protein, e.g., from a natural source. Typically, it will be useful in isolating a gene from human, and similar procedures will be applied to isolate genes from other species, e.g., warm blooded animals, such as birds and mammals. Cross hybridization will allow isolation of

analogous factors from other species. A number of different approaches should be available to successfully isolate a suitable nucleic acid clone.

The purified protein or defined peptides are useful for generating antibodies by standard methods, as described above. Synthetic peptides or purified protein can be presented to an immune system to generate monoclonal or polyclonal antibodies. See, e.g.: Coligan (1991) Current Protocols in Immunology Wiley/Greene; and Harlow and Lane (1989) Antibodies: A Laboratory Manual Cold Spring Harbor Press. Alternatively, the AP-1 complex can be used as a specific binding reagent, and advantage can be taken of its specificity of binding, very much as an antibody would be used.

For example, a specific binding composition will be useful for screening an expression library made from a cell line which expresses an NF-AT120 protein. The screening can be standard staining of surface-expressed protein, or panning. Screening of intracellular expression can also be performed by various staining or immunofluorescence procedures. The binding compositions could be used to affinity purify or sort out cells expressing the protein.

The peptide segments can also be used to predict appropriate oligonucleotides to screen a library. The genetic code can be used to select appropriate oligonucleotides useful as probes for screening; see, e.g., SEQ ID NO: 6 through 24. In combination with polymerase chain reaction (PCR) techniques, synthetic oligonucleotides were useful in selecting correct clones from a library, e.g., inserts with sequences described in SEQ ID NO: 34, 36, 38 or 40. Complementary sequences will also be used as probes, primers, or antisense molecules. It is recognized that minimally degenerate primers will typically be preferred, so long as the primers are of sufficient length, e.g., preferably at least about 18 nucleotides.

This invention contemplates use of isolated DNA or fragments to encode a biologically active corresponding NF-AT120 protein polypeptide. In addition, this invention covers isolated or recombinant DNA which encodes a biologically active protein or polypeptide which is capable of hybridizing under appropriate conditions with the DNA sequences described herein. Said biologically active protein or polypeptide can be an intact factor, or fragment, and have an amino acid sequence as disclosed in SEQ ID NO: 1 through 5, 35, 37, 39 or 41. Further, this invention covers the use of isolated or recombinant DNA, or fragments thereof, which encode proteins that are homologous to an NF-AT120 protein or which was isolated using cDNA encoding an NF-AT120 protein as a probe, e.g., other members of the subfamilies of NF-AT120. The isolated DNA

can have the respective regulatory sequences in the 5'- and 3'-flanks, e.g., promoters, enhancers, poly-A addition signals, and others.

The phosphoramidite method described by Beaucage and Carruthers (1981) Tetra. Letts. 22:1859-1862, will produce suitable synthetic DNA fragments. A double-stranded fragment will often be obtained either by synthesizing the complementary strand and annealing the strand together under appropriate conditions or by adding the complementary strand using DNA polymerase with an appropriate primer sequence, e.g., by PCR techniques.

An "isolated" nucleic acid is a nucleic acid, e.g., an RNA, DNA, or a mixed polymer, which is substantially separated from other components that naturally accompany a native sequence, e.g., ribosomes, polymerases and flanking genomic sequences from the originating species. The term embraces a nucleic acid sequence which has been removed from its naturally occurring environment, and includes recombinant or cloned DNA isolates and chemically synthesized analogues or analogues biologically synthesized by heterologous systems. A substantially pure molecule includes isolated forms of the molecule.

An isolated nucleic acid will generally be a homogeneous composition of molecules, but will, in some embodiments, contain minor heterogeneity. This heterogeneity is typically found at the polymer ends or portions not critical to a desired biological function or activity.

A "recombinant" nucleic acid is defined either by its method of production or by its structure. In reference to its method of production, e.g., a process for making a product, the process is use of recombinant nucleic acid techniques, e.g., involving human intervention in the nucleotide sequence, typically selection or production, and generally using some *in vitro* steps. In reference to its structure, it can be a nucleic acid made by fusion of two fragments which are not naturally contiguous to each other, but it is meant to exclude products of nature, e.g., naturally occurring mutants. Thus, for example, products made by transforming cells with any unnaturally occurring vector are encompassed, as are nucleic acids comprising sequences derived using any synthetic oligonucleotide process. Such a process is often used to replace a codon with a redundant codon encoding the same or a conservative amino acid, while typically introducing or removing a sequence-recognition site.

Alternatively, it is performed to join together nucleic acid segments of desired functions to generate a single genetic entity comprising a desired combination of functions not found in the commonly available natural forms. Restriction enzyme recognition sites are often the target of such artificial

manipulations, but other site-specific targets, e.g., promoters, DNA replication sites, regulation sequences, control sequences, or other useful features, may be incorporated by design. A similar concept is intended for a recombinant polypeptide, e.g., a fusion polypeptide. Specifically included are synthetic nucleic acids which, by genetic code redundancy, encode polypeptides similar to fragments of these antigens, and fusions of sequences from various different species variants.

A significant "fragment" in a nucleic acid context is a contiguous segment of at least about 17 nucleotides nucleotides, generally at least 23 or 29 nucleotides, often at least 35 or 41 nucleotides, preferably at least 47 or 53 nucleotides, and in particularly preferred embodiments will be of 56 or more nucleotides.

A DNA which codes for an NF-AT120 protein will be particularly useful to identify genes, mRNA, and cDNA species which code for related or homologous proteins, e.g., the members of the C, P and X subfamilies, as well as DNAs which code for homologous proteins from different species. There are likely homologues in other orders and species, including primates and rodents. Various NF-AT120 proteins should be homologous and are encompassed herein. However, even proteins that have a more distant evolutionary relationship to the NF-AT120 protein can readily be isolated under appropriate conditions using these sequences if they are sufficiently homologous. Primate NF-AT120 proteins are of particular interest.

This invention further covers recombinant DNA molecules and fragments having a DNA sequence identical to or highly homologous to the isolated DNAs set forth herein. In particular, the sequences will often be operably linked to DNA segments which control transcription, translation, and DNA replication. Alternatively, recombinant clones derived from the genomic sequences, e.g., containing introns, will be useful for transgenic studies, including, e.g., transgenic cells and organisms, and for gene therapy. See, e.g.: Goodnow (1992) "Transgenic Animals" in Roitt (ed.) Encyclopedia of Immunology Academic Press, San Diego, pp. 1502-1504; Travis (1992) Science 256:1392-1394; Kuhn et al. (1991) Science 254:707-710; Capecchi (1989) Science 244:1288; Robertson (1987) (ed.) Teratocarcinomas and Embryonic Stem Cells: A Practical Approach IRL Press, Oxford; and Rosenberg (1992) J. Clinical Oncology 10:180-199.

Homologous nucleic acid sequences, when compared, exhibit significant similarity. The standards for homology in nucleic acids either are measures for

homology generally used in the art by sequence comparison or are based upon hybridization conditions. The hybridization conditions are described in greater detail below.

Substantial homology in the context of comparing nucleic acid sequences means that the segments are identical when optimally aligned, allowing for appropriate nucleotide insertions or deletions, in at least about 50% or even 59% of the nucleotides, generally at least 65% or even 71%, usually at least about 77% or even 85%, preferably at least about 95 to 98% or more, and, in particular embodiments, as many as about 99% or more of the nucleotides. Since dsDNA uses specific nucleotide pairing, the comparison can also be made on the basis of a complementary strand.

Alternatively, substantial homology exists when the segments will hybridize under selective hybridization conditions to a strand or its complement, typically using a sequence derived from SEQ ID NO: 6 through 24, 34, 36, 38 and 40. Typically, selective hybridization will occur when there is at least about 55% homology over a stretch of at least about 40 nucleotides, preferably at least about 75% over a stretch of at least about 25 nucleotides, and most preferably at least about 90%. See Kanehisa (1984) Nucl. Acids Res., 12:203-213. The length of homology comparison, as described, may be over longer stretches, and in certain embodiments will be over a stretch of at least about 17 nucleotides, usually at least about 24 nucleotides, typically at least about 40 nucleotides, and preferably at least about 75 to 100 or more nucleotides.

Stringent conditions, in referring to homology in the hybridization context, will be stringent combined conditions of salt, temperature, organic solvents, and other variables, typically those controlled in hybridization reactions. Stringent temperature conditions will usually include temperatures in excess of about 30°C, usually in excess of about 37°C, typically in excess of about 55°C, and preferably in excess of about 70°C. Stringent salt conditions will ordinarily be less than about 1000 mM, usually less than about 400 mM, typically less than about 300 mM, and preferably less than about 150 mM. However, the combination of variables is much more important than the measure of any single variable: see, e.g., Wetmur and Davidson (1968) J. Mol. Biol. 31:349-370.

NF-AT120 protein from other mammalian species can be cloned and isolated by cross-species hybridization of closely related species. Homology may be relatively low between distantly related species, and thus hybridization of relatively closely related species is advisable. Alternatively, preparation of an

antibody which exhibits less species specificity may be useful in expression cloning approaches.

Making NF-AT120 protein: Mimetics

DNA which encodes an NF-AT120 protein or fragments thereof can be obtained by chemical synthesis, by screening cDNA libraries, or by screening genomic libraries prepared from a wide variety of cell lines or tissue samples.

This DNA can be expressed in a wide variety of host cells for the synthesis of a full-length protein or fragments which can in turn, for example, be used to generate polyclonal or monoclonal antibodies; for binding studies; for construction and expression of modified molecules; and for structure/function studies. Each antigen or its fragments can be expressed in host cells that are transformed or transfected with appropriate expression vectors. These molecules can be substantially purified to be free of protein or cellular contaminants, other than those derived from the recombinant host, and therefore are particularly useful in pharmaceutical compositions when combined with a pharmaceutically acceptable carrier and/or diluent. The antigen, or portions thereof, may be expressed as fusions with other proteins.

Expression vectors are typically self-replicating DNA or RNA constructs containing the desired antigen gene or its fragments, usually operably linked to suitable genetic control elements that are recognized in a suitable host cell. These control elements are capable of effecting expression within a suitable host. The specific type of control elements necessary to effect expression will depend upon the eventual host cell used. Generally, the genetic control elements can include a prokaryotic promoter system or a eukaryotic promoter expression control system, and typically include a transcriptional promoter, an optional operator to control the onset of transcription, transcription enhancers to elevate the level of mRNA expression, a sequence that encodes a suitable ribosome binding site, and sequences that terminate transcription and translation. Expression vectors also usually contain an origin of replication that allows the vector to replicate independently of the host cell.

The vectors of this invention contain DNA which encodes an NF-AT120 protein, or a fragment thereof, typically encoding a biologically active polypeptide. The DNA can be under the control of a viral promoter and can encode a selection marker. This invention further contemplates use of such expression vectors which are capable of expressing eukaryotic cDNA coding for an NF-AT120 protein in a prokaryotic or eukaryotic host, where the vector is

compatible with the host and where the eukaryotic cDNA coding for the protein is inserted into the vector such that growth of the host containing the vector expresses the cDNA in question. Usually, expression vectors are designed for stable replication in their host cells or for amplification to greatly increase the total number of copies of the desirable gene per cell. It is not always necessary to require that an expression vector replicate in a host cell: e.g., it is possible to effect transient expression of the protein or its fragments in various hosts using vectors that do not contain an origin of replication that is recognized by the host cell. It is also possible to use vectors that cause integration of an NF-AT120 protein gene or its fragments into the host DNA by recombination, or to integrate a promoter which controls expression of an endogenous gene.

Vectors, as used herein, comprise plasmids, viruses, bacteriophage, integratable DNA fragments, and other vehicles which enable the integration of DNA fragments into the genome of the host. Expression vectors are specialized vectors which contain genetic control elements that effect expression of operably linked genes. Plasmids are the most commonly used form of vector, but all other forms of vectors which serve an equivalent function and which are, or become, known in the art are suitable for use herein. See, e.g., Pouwels et al. (1985 and Supplements) Cloning Vectors: A Laboratory Manual, Elsevier, N.Y., and Rodriquez et al. (1988) (eds.) Vectors: A Survey of Molecular Cloning Vectors and Their Uses, Buttersworth, Boston, MA.

Transformed cells include cells, preferably mammalian, that have been transformed or transfected with vectors that contain the NF-AT120 protein gene and have been constructed using recombinant DNA techniques. Transformed host cells usually express the protein or its fragments; however, for purposes of cloning, amplifying, and manipulating its DNA, they do not need to express the protein. This invention further contemplates culturing transformed cells in a nutrient medium, thus permitting the protein to accumulate in the culture. The protein can be recovered, either from the culture or from the culture medium.

For purposes of this invention, DNA sequences are operably linked when they are functionally related to each other. For example, DNA for a presequence or secretory leader is operably linked to a polypeptide if it is expressed as a preprotein or participates in directing the polypeptide to the cell membrane or in secretion of the polypeptide. A promoter is operably linked to a coding sequence if it controls the transcription of the polypeptide; a ribosome binding site is operably linked to a coding sequence if it is positioned to permit translation. Usually, "operably linked" means contiguous and in reading frame;

however, certain genetic elements such as repressor genes are not contiguously linked but still bind to operator sequences that in turn control expression.

Suitable host cells include prokaryotes, lower eukaryotes, and higher eukaryotes. Prokaryotes include both gram negative and gram positive organisms, e.g., *E. coli* and *B. subtilis*. Lower eukaryotes include yeasts, e.g., *S. cerevisiae* and *Pichia*, and species of the genus *Dictyostelium*. Higher eukaryotes include established tissue culture cell lines from animal cells, both of non-mammalian origin, e.g., insect cells, and birds, and of mammalian origin, e.g., human, primates, and rodents.

Prokaryotic host-vector systems include a wide variety of vectors for many different species. As used herein, *E. coli* and its vectors will be used generically to include equivalent prokaryotes and vectors used therein . A representative vector for amplifying DNA is pBR322 or many of its derivatives. Vectors that can be used to express the NF-AT120 proteins or its fragments include, but are not limited to, such vectors as: those containing the lac promoter (pUC-series); trp promoter (pBR322-trp); lpp promoter (the pIN-series); lambda-pP or pR promoters (pOTS); or hybrid promoters such as ptac (pDR540). See Brosius et al. (1988) "Expression Vectors Employing Lambda-, trp-, lac-, and lpp-derived Promoters", in Rodriguez and Denhardt (eds.) Vectors: A Survey of Molecular Cloning Vectors and Their Uses, Buttersworth, Boston, Chapter 10, pp. 205-236; and Balbas and Bolivar (1990) Methods in Enzymology, 185:14-37.

Lower eukaryotes, e.g., yeasts and *Dictyostelium*, may be transformed with NF-AT120 protein sequence containing vectors. For purposes of this invention, the most common lower eukaryotic host is the baker's yeast, *Saccharomyces cerevisiae*. It will be used to generically represent lower eukaryotes, although a number of other strains and species are also available. Yeast vectors typically consist of a replication origin (unless they are of the integrating type), a selection gene, a promoter, DNA encoding the desired protein or its fragments, and sequences for termination of translation, polyadenylation, and termination of transcription. Suitable expression vectors for yeast include such constitutive promoters as 3-phosphoglycerate kinase and various other glycolytic enzyme gene promoters or such inducible promoters as the alcohol dehydrogenase-2 promoter or metallothionine promoter: see, e.g., Stearns et al. (1990) Methods in Enzymology 185:280-297; and chapter 29 in Wu et al. (eds.) (1989) Recombinant DNA Methodology, Academic Press, San Diego.

Higher eukaryotic tissue culture cells are the preferred host cells for expression of the functionally active NF-AT120 protein. In principle, any higher eukaryotic tissue culture cell line is workable, e.g., insect baculovirus expression systems, whether from an invertebrate or vertebrate source: see, e.g., Miller (1988) Ann. Rev. Microbiol. 42:177-99. However, mammalian cells are generally preferred for their protein processing patterns, both cotranslational and posttranslational. Transformation or transfection and propagation of such cells is described in, e.g., Ausubel, et al. (eds.) (1987 and Supplements) Current Protocols in Molecular Biology, Greene/Wiley, New York.

Examples of useful cell lines include HeLa cells, Chinese hamster ovary (CHO) cell lines, baby rat kidney (BRK) cell lines, insect cell lines, bird cell lines, and monkey (COS) cell lines. Expression vectors for such cell lines usually include an origin of replication, a promoter, a site for initiation of translation, RNA splice sites (if genomic DNA is used), a polyadenylation site, and a site for termination of transcription. These vectors also usually contain a selection gene or amplification gene.

Suitable expression vectors may be plasmids, viruses, or retroviruses carrying promoters derived, e.g., from such sources as adenovirus, SV40, parvoviruses, vaccinia virus, or cytomegalovirus. Representative examples of suitable expression vectors include pCDNA1; pCD [see Okayama et al. (1985) Mol. Cell Biol., 5:1136-1142]; pMC1neo Poly-A [see Thomas et al. (1987) Cell, 51:503-512]; and a baculovirus vector such as pAC 373 or pAC 610.

It will often be desired to express an NF-AT120 protein polypeptide in a system which provides a specific or defined glycosylation pattern. In this case, the usual pattern will be that provided naturally by the expression system. However, the pattern will be modifiable by exposing the polypeptide, e.g., an unglycosylated form, to appropriate glycosylating proteins introduced into a heterologous expression system. For example, the NF-AT120 protein gene may be co-transformed with one or more genes encoding mammalian or other glycosylating enzymes. Using this approach, certain mammalian glycosylation patterns will be achievable or approximated in prokaryote or other cells.

The NF-AT120 protein, or a fragment thereof, may be engineered to be linked by phosphatidylinositol (PI) to a cell membrane, but can be removed from membranes by treatment with a phosphatidylinositol-cleaving enzyme, e.g., phosphatidylinositol phospholipase-C. This releases the antigen in a biologically active form, and allows purification by standard procedures of protein chemistry. See, e.g.: Low (1989) Biochim. Biophys. Acta, 988:427-454;

Tse et al. (1985) Science, 230:1003-1008; and Brunner et al. (1991) J. Cell Biol., 114:1275-1283.

Now that NF-AT120 proteins have been characterized, fragments or derivatives thereof can be prepared by conventional processes for synthesizing peptides. These include processes such as are described by: Stewart and Young in Solid Phase Peptide Synthesis (1984), Pierce Chemical Co., Rockford, IL; Bodanszky and Bodanszky, The Practice of Peptide Synthesis, (1984), Springer-Verlag, New York; and Bodanszky, The Principles of Peptide Synthesis, (1984), Springer-Verlag, New York. For example, an azide process, an acid chloride process, an acid anhydride process, a mixed anhydride process, an active ester process (for example, 4-nitrophenyl ester, N-hydroxy-succinimide ester, or cyanomethyl ester), a carbodiimidazole process, an oxidative-reductive process, or a dicyclohexylcarbodiimide/additive process, can be used. Solid-phase and solution-phase syntheses are both applicable to the foregoing processes.

The NF-AT120 protein, fragments, or derivatives are suitably prepared in accordance with the above processes as typically employed in peptide synthesis, generally either by a so-called stepwise process which comprises condensing an amino acid to the terminal amino acid, one by one in sequence, or by coupling peptide fragments to the terminal amino acid. Amino groups that are not being used in the coupling reaction are typically protected to prevent coupling at an incorrect location.

If a solid phase synthesis is adopted, the C-terminal amino acid is bound to an insoluble carrier or support through its carboxyl group. The insoluble carrier is not particularly limited as long as it has a binding capability to a reactive carboxyl group. Examples of such insoluble carriers include halomethyl resins, such as chloromethyl resin or bromomethyl resin, hydroxymethyl resins, phenol resins, tert-alkyloxycarbonyl-hydrazide resins, and the like.

An amino group-protected amino acid is bound in sequence through condensation of its activated carboxyl group and the reactive amino group of the previously formed peptide or chain, to synthesize the peptide step-by-step. After the complete sequence has been synthesized , the peptide is split off from the insoluble carrier to produce the peptide. This solid-phase approach is generally described by Merrifield et al. (1963) in J. Am. Chem. Soc., 85:2149-2156.

The prepared protein and fragments thereof can be isolated and purified from the reaction mixture by means of peptide separation, for example, by extraction, precipitation, electrophoresis and various forms of chromatography,

and the like. The NF-AT120 proteins of this invention can be obtained in varying degrees of purity depending upon their desired use. Purification can be accomplished by use of the protein purification techniques disclosed herein or by the use of the antibodies herein described in immunoabsorbant affinity chromatography. This immunoabsorbant affinity chromatography is carried out by first linking the antibodies to a solid support and then contacting the linked antibodies with solubilized lysates of appropriate source cells, lysates of other cells expressing the protein, or lysates or supernatants of cells producing the NF-AT120 protein as a result of DNA techniques; see below.

Uses

The present invention provides reagents which will find use in diagnostic applications as described elsewhere herein, e.g., in the general description for developmental abnormalities, or below in the description of kits for diagnosis.

This invention also provides reagents with significant therapeutic value. NF-AT120 proteins (naturally occurring or recombinant), fragments thereof and antibodies thereto, along with compounds identified as having binding affinity to NF-AT120 protein, should be useful in the treatment of conditions associated with abnormal physiology or development, including abnormal proliferation, e.g., cancerous conditions, or degenerative conditions. In particular, modulation of expression of cytokine genes is possible, but an NF-AT120 may be a component in regulatory proteins affecting other genes. Abnormal proliferation, regeneration, degeneration, and atrophy may be modulated by appropriate therapeutic treatment using the compositions provided herein. For example, a disease or disorder associated with abnormal transcriptional regulation by an NF-AT120 protein should be a likely target for an inhibitor or stimulator. The factor probably plays a role in regulation or development of hematopoietic cells, e.g., lymphoid cells, which affect immunological responses, e.g., autoimmune or immunodeficiency disorders.

Antibodies raised to recombinant NF-AT120 protein can be prepared and purified and then administered to a patient. These reagents can be combined for therapeutic use with additional active or inert ingredients, e.g., in conventional pharmaceutically acceptable carriers or diluents, e.g., immunogenic adjuvants, along with physiologically innocuous stabilizers and excipients. These combinations can be filtered sterile and placed into dosage forms by lyophilization in dosage vials or storage in stabilized aqueous

preparations. This invention also contemplates use of antibodies or binding fragments thereof, including forms which do not bind complement.

Drug screening using NF-AT120 or fragments thereof can be performed to identify compounds having binding affinity to NF-AT120 protein, including isolating associated components. Similar screening for compounds which interact with AP-1 will identify proteins which modulate NF-AT function. Subsequent biological assays can then be utilized to determine if the compound has intrinsic modulatory activity and is therefore useful in that it blocks the activity of the protein. Likewise, a compound having intrinsic stimulating activity can activate the receptor and is thus an agonist in that it simulates the activity of NF-AT120 protein. This invention further contemplates the therapeutic use of antibodies to NF-AT120 protein as transcription modulators. This approach should be particularly useful with other NF-AT120 protein species variants.

The quantities of reagents necessary for effective therapy will depend upon many different factors, including means of administration, target site, physiological state of the patient, and other medicants administered. Thus, treatment dosages should be titrated to optimize safety and efficacy. Typically, dosages used *in vitro* may provide useful guidance in the amounts useful for *in situ* administration of these reagents. Animal testing of effective doses for treatment of particular disorders will provide further predictive indication of human dosage. Various considerations are described, e.g., in Gilman et al. (eds.) (1990) Goodman and Gilman's: The Pharmacological Bases of Therapeutics, 8th Ed., Pergamon Press; and Remington's Pharmaceutical Sciences, 17th ed. (1990), Mack Publishing Co., Easton, Penn. Methods for administration are discussed therein and below, e.g., oral, intravenous, intraperitoneal, or intramuscular administration, transdermal diffusion, and others. Pharmaceutically acceptable carriers will include water, saline, buffers, and other compounds described, e.g., in The Merck Index, Merck & Co., Rahway, New Jersey. Dosage ranges would ordinarily be expected to be in amounts lower than 1 mM concentrations, typically less than about 10 µM concentrations, usually less than about 100 nM, preferably less than about 10 pM (picomolar), and most preferably less than about 1 fM (femtomolar), with an appropriate carrier. Slow-release formulations or a slow-release apparatus will often be utilized for continuous administration; see also Langer, (1990) Science 249:1527-1533.

NF-AT120 protein, fragments thereof, and antibodies to it or its fragments may be administered directly to the host to be treated. Alternatively, depending

on the size of the compounds, it may be desirable to conjugate them to carrier proteins such as ovalbumin or serum albumin prior to their administration. Therapeutic formulations may be administered in many conventional dosage formulations.

Whereas it is possible for the active ingredient to be administered alone, it is preferable to present it as a pharmaceutical formulation. Formulations typically comprise at least one active ingredient, as defined above, together with one or more acceptable carriers thereof. Each carrier should be both pharmaceutically and physiologically acceptable in the sense of being compatible with the other ingredients and not injurious to the patient. Formulations include those suitable for oral, rectal, nasal, or parenteral administration (including subcutaneous, intramuscular, intravenous and intradermal administration).

The formulations may conveniently be presented in unit dosage form and may be prepared by any methods well known in the art of pharmacy. See, e.g.: Gilman et al. (eds.) (1990) Goodman and Gilman's: The Pharmacological Bases of Therapeutics, 8th Ed., Pergamon Press; and Remington's Pharmaceutical Sciences, 17th ed. (1990), Mack Publishing Co., Easton, Penn.; Avis et al. (eds.) (1993) Pharmaceutical Dosage Forms: Parenteral Medications, Dekker, New York; Lieberman et al. (eds.) (1990) Pharmaceutical Dosage Forms: Tablets, Dekker, New York; and Lieberman et al. (eds.) (1990) Pharmaceutical Dosage Forms: Disperse Systems, Dekker, New York. The therapy of this invention may be combined with or used in association with other chemotherapeutic or chemopreventive agents.

Both the naturally occurring and the recombinant form of the NF-AT120 proteins of this invention are particularly useful in kits and assay methods which are capable of screening compounds for binding activity to the proteins, or the AP-1 complex or DNA. Several methods of automating assays have been developed in recent years so as to permit screening of tens of thousands of compounds in a short period. See, e.g., Fodor et al., (1991) Science 251:767-773, which describes means for testing of binding affinity by a plurality of defined polymers synthesized on a solid substrate. Structural diversity libraries can be used. The development of suitable assays can be greatly facilitated by the availability of large amounts of purified, soluble NF-AT120 protein as provided by this invention.

For example, modulators can normally be found once the protein or complex has been structurally defined. Testing of potential protein analogues is now possible with the development of highly automated assay methods using

the information provided herein. In particular, new agonists and antagonists will be discovered by using screening techniques described herein. Of particular importance are compounds found to have a combined binding affinity for multiple AP-1 types, e.g., compounds which can serve as modulators for species variants of NF-AT120 protein.

This invention is particularly useful for screening compounds by using recombinant NFAT120 in a variety of drug-screening techniques. The advantages of using a recombinant protein in screening for specific proteins include: (a) improved renewable source of the protein from a specific source; (b) potentially greater number of NF-AT120 proteins per cell giving better signal-to-noise ratio in assays; and (c) species variant specificity (theoretically giving greater biological and disease specificity).

One method of drug screening utilizes eukaryotic or prokaryotic host cells which are stably transformed with recombinant DNA molecules expressing the protein. Cells may be isolated which express the protein in isolation from others. Such cells, either in viable or fixed form, can be used for standard binding assays. See also: Parce et al. (1989) Science, 246:243-247; and Owicki et al. (1990) Proc. Nat'l Acad. Sci. USA, 87:4007-4011; which describe sensitive methods to detect cellular responses. Competitive assays are particularly useful, where the cells (source of NF-AT120 protein) are contacted and incubated with a labeled receptor or antibody having known binding affinity to the protein, such as ^{125}I-antibody, and a test sample whose binding affinity to the binding composition is being measured. The bound and the free labeled binding compositions are then separated to assess the degree of protein binding. The amount of test compound bound is inversely proportional to the amount of labeled receptor binding to the known source. Any one of numerous techniques can be used to separate bound from free protein to assess the degree of protein binding. This separation step could typically involve a procedure such as adhesion to filters followed by washing, adhesion to plastic followed by washing, or centrifugation of the cell membranes. Viable cells could also be used to screen for the effects of drugs on NF-AT120 protein mediated functions: e.g., second messenger levels, i.e., Ca^{2+}; cell proliferation; inositol phosphate pool changes; and others. Some detection methods allow for elimination of a separation step, e.g., a proximity sensitive detection system. Calcium sensitive dyes will be useful for detecting Ca^{2+} levels, with a fluorimeter or a fluorescence cell sorting apparatus.

Another method utilizes membranes from transformed eukaryotic or prokaryotic host cells as the source of the NF-AT120 protein. These cells are stably transformed with DNA vectors directing the expression of an NF-AT120 protein, e.g., an engineered membrane bound form. Essentially, the membranes would be prepared from the cells and used in a binding assay such as the competitive assay set forth above.

Still another approach is to use solubilized and unpurified NF-AT120 protein or solubilized and purified NF-AT120 protein from transformed eukaryotic or prokaryotic host cells. This allows for a "molecular" binding assay with the advantages of increased specificity, the ability to automate, and high drug test throughput.

Another technique for drug screening involves an approach which provides high throughput screening for compounds having suitable binding affinity to NF-AT120 and is described in detail International Patent Application no. WO 84/03564 (Commonwealth Serum Labs.), published on September 13 1984. First, large numbers of different small peptide test compounds are synthesized on a solid substrate, e.g., plastic pins or some other appropriate surface; see Fodor et al. (1991). Then all the pins are reacted with solubilized and unpurified NF-AT120, or with solubilized and purified NF-AT120, and washed. The next step involves detecting bound NF-AT120.

Rational drug design may also be based upon structural studies of the molecular shapes of the NF-AT120 protein and other binding partners. These may be other proteins which mediate other functions in response to NF-AT120 binding, or other proteins which normally interact with the factor, or even DNA segments. One means for determining which sites interact with specific other proteins is a physical structure determination, e.g., x-ray crystallography or two-dimensional NMR techniques. These will provide guidance as to which amino acid residues form molecular contact regions. For a detailed description of protein structural determination, see, e.g., Blundell and Johnson (1976) Protein Crystallography, Academic Press, New York.

Purified NF-AT120 protein can be coated directly onto plates for use in the aforementioned drug-screening techniques. However, non-neutralizing antibodies to these factors can be used as capture antibodies to immobilize the respective factor on the solid phase.

Kits

This invention also contemplates use of NF-AT120 proteins, fragments thereof, peptides, and their fusion products in a variety of diagnostic kits and methods for detecting the presence of an NF-AT120. Typically the kit will have a compartment containing either a defined NF-AT120 protein peptide or gene segment or a reagent which recognizes one or the other, e.g., antibodies or fragments thereof.

A kit for determining the binding affinity of a test compound to an NF-AT120 protein would typically comprise: a test compound; a labeled compound, for example an antibody having known binding affinity for the protein; a source of NF-AT120 protein (naturally occurring or recombinant); and a means for separating bound from free labeled compound, such as a solid phase for immobilizing the protein. Once compounds are screened, those having suitable binding affinity to the protein can be evaluated in suitable biological assays to determine whether they act as modulators of transcription. The availability of recombinant NF-AT120 polypeptides also provide well-defined standards for calibrating such assays.

A preferred kit for determining the concentration of, for example, an NF-AT120 protein in a sample would typically comprise a labeled compound, e.g., AP-1 or antibody, having known binding affinity for the protein, a source of NF-AT120 (naturally occurring or recombinant) and a means for separating the bound from free labeled compound, for example, a solid phase for immobilizing the NF-AT120 protein. Compartments containing reagents, and instructions, will normally be provided.

Antibodies, including antigen-binding fragments, specific for the NF-AT120 protein or fragments are useful in diagnostic applications to detect the presence of elevated levels of NF-AT120 protein and/or its fragments. Such diagnostic assays can employ lysates, live cells, fixed cells, immuno-fluorescence, cell cultures, and body fluids, and further can involve the detection of antigens related to the NF-AT120 in serum, or the like. Diagnostic assays may be homogeneous (without a separation step between free reagent and bound complex) or heterogeneous (with a separation step). Various commercial assays exist, such as radioimmunoassay (RIA), enzyme-linked immunosorbent assay (ELISA), enzyme immunoassay (EIA), enzyme-multiplied immunoassay technique (EMIT), substrate-labeled fluorescent immunoassay (SLFIA), and the like. For example, unlabeled antibodies can be employed by using a second antibody which is labeled and which recognizes the antibody to an NF-AT120

protein or to a particular fragment thereof. Similar assays have also been extensively discussed in the literature. See, e.g., Harlow and Lane (1988) Antibodies: A Laboratory Manual, CSH.

Anti-idiotypic antibodies may have similar use to diagnose presence of antibodies against an NF-AT120 protein, since the latter antibodies may be diagnostic of various abnormal states, e.g., autoimmune conditions. For example, overproduction of NF-AT120 protein may result in production of various immunological reactions which may be diagnostic of abnormal physiological states, particularly in proliferative cell conditions such as cancer or abnormal differentiation.

Frequently, the reagents for diagnostic assays are supplied in kits, so as to optimize the sensitivity of the assay. For the subject invention, depending upon the nature of the assay, the protocol, and the label, either labeled or unlabeled antibody or labeled NF-AT120 protein is provided. This is usually in conjunction with other additives, such as buffers, stabilizers, materials necessary for signal production such as substrates for enzymes, and the like. Preferably, the kit will also contain instructions for proper use and disposal of the contents after use. Typically the kit has compartments for each useful reagent. Desirably, the reagents are provided as a dry lyophilized powder, and may be reconstituted in an aqueous medium providing appropriate concentrations of reagents for performing the assay.

Any of the aforementioned constituents of the drug-screening assay and the diagnostic assay may be used without modification or may be modified in a variety of ways. For example, labeling may be achieved by covalently or non-covalently joining a moiety which directly or indirectly provides a detectable signal. In any of these assays, the test compound, NF-AT120 protein, or antibodies thereto can be labeled either directly or indirectly. Possibilities for direct labeling include label groups: radiolabels such as ^{125}I, enzymes (U.S. Pat. No. 3,645,090) such as peroxidase and alkaline phosphatase, and fluorescent labels (U.S. Pat. No. 3,940,475) capable of monitoring the change in fluorescence intensity, wavelength shift, or fluorescence polarization. Possibilities for indirect labeling include biotinylation of one constituent followed by binding to avidin coupled to one of the above label groups.

There are also numerous methods of separating the bound from the free antibody, or alternatively the bound from the free test compound. The NF-AT120 protein can be immobilized on various matrixes followed by washing. Suitable matrixes include plastic such as an ELISA plate, filters, and beads. Methods of

immobilizing the NF-AT120 protein to a matrix include, without limitation, direct adhesion to plastic, use of a capture antibody, chemical coupling, and biotin-avidin. The last step in this approach generally involves the precipitation of protein/binding partner or protein/antibody complex by various methods including those utilizing, e.g., an organic solvent such as polyethylene glycol or a salt such as ammonium sulfate. Other suitable separation techniques include, without limitation, the fluorescein antibody magnetizable particle method described in Rattle et al. (1984) Clin. Chem. 30:1457-1461, and the double antibody magnetic particle separation as described in U.S. Pat. No. 4,659,678.

Methods for linking proteins or their fragments to the various labels have been extensively reported in the literature. Many of the techniques involve the use of activated carboxyl groups either through the use of carbodiimide or active esters to form peptide bonds, the formation of thioethers by reaction of a mercapto group with an activated halogen such as chloroacetyl, or an activated olefin such as maleimide, for linkage, or the like. Fusion proteins will also find use in these applications.

Another diagnostic aspect of this invention involves use of oligonucleotide or polynucleotide sequences taken from the sequence of an NF-AT120 protein. These sequences can be used as probes for detecting levels of the message in samples from patients suspected of having an abnormal condition, e.g., cancer or developmental problem. The preparation of both RNA and DNA nucleotide sequences, the labeling of the sequences, and the preferred size of the sequences has been described above. Normally an oligonucleotide probe should have at least about 14 nucleotides and usually at least about 18 nucleotides, and the polynucleotide probes may be up to several kilobases. Various labels may be employed, most commonly radionuclides, particularly ^{32}P.

Other techniques may also be employed, such as using biotin modified nucleotides for introduction into a polynucleotide. The biotin then serves as the site for binding to avidin or antibodies, which may be labeled with a wide variety of labels, such as radionuclides, fluorescers, enzymes, or the like. Alternatively, antibodies may be employed which can recognize specific duplexes, including DNA duplexes, RNA duplexes, DNA-RNA hybrid duplexes, or DNA-protein duplexes. The antibodies in turn may be labeled and the assay carried out where the duplex is bound to a surface, so that upon the formation of duplex on the surface, the presence of antibody bound to the duplex can be detected. The use of probes to the novel anti-sense RNA may be carried out in any conventional techniques such as nucleic acid hybridization, plus and minus

screening, recombinational probing, hybrid released translation (HRT), and hybrid arrested translation (HART). This also includes amplification techniques such as polymerase chain reaction (PCR).

Diagnostic kits which also test for the qualitative or quantitative presence of other markers are also contemplated. Diagnosis or prognosis may depend on the combination of multiple indications used as markers. Thus, kits may test for combinations of markers; see, e.g., Viallet et al. (1989) Progress in Growth Factor Res., 1:89-97.

EXAMPLES

The broad scope of this invention is best understood with reference to the following Examples, which are intended to illustrate and not to limit the invention to specific embodiments. In particular, the selected vectors and hosts, the concentration of reagents, the temperatures, and the values of other variables are only to exemplify the application of the present invention and are not to be considered limitations thereof.

General Methods

Some of the standard methods are described or referenced, e.g., in Maniatis et al. (1982) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor Press; Sambrook et al. (1989) Molecular Cloning: A Laboratory Manual, (2d ed.), vols 1-3, CSH Press, NY; Ausubel et al., Biology, Greene Publishing Associates, Brooklyn, NY; Ausubel et al. (1987 and Supplements) Current Protocols in Molecular Biology, Greene/Wiley, New York; Innis et al. (eds.) (1990) PCR Protocols: A Guide to Methods and Applications, Academic Press, N.Y.; and Coligan, et al. (1991 and Supplements) Current Protocols in Immunology, Greene/Wiley, N.Y.

Methods for protein purification include such methods as precipitation with ammonium sulfate, column chromatography, electrophoresis, centrifugation, crystallization, and others. See, e.g.: Ausubel et al. (1987 and periodic supplements); Deutscher (1990) "Guide to Protein Purification" in Methods in Enzymology, vol. 182, and other volumes in this series; and manufacturers' literature on use of protein purification products, e.g., Pharmacia, Piscataway, N.J., or Bio-Rad, Richmond, CA. Combination with recombinant techniques allows fusion to appropriate segments, e.g., to a FLAG sequence or an equivalent which can be fused through a protease-removable sequence. See,

e.g.: Hochuli (1989) Chemische Industrie, 12:69-70; Hochuli (1990) "Purification of Recombinant Proteins with Metal Chelate Absorbent" in Setlow (ed.) Genetic Engineering, Principle and Methods, 12:87-98, Plenum Press, N.Y.; and Crowe et al. (1992) QIAexpress: The High Level Expression & Protein Purification System, QUIAGEN, Inc., Chatsworth, CA.

FACS analyses are described in: Melamed et al. (1990) Flow Cytometry and Sorting, Wiley-Liss, Inc., New York, NY; Shapiro (1988) Practical Flow Cytometry, Liss, New York, NY; and Robinson et al. (1993) Handbook of Flow Cytometry Methods, Wiley-Liss, New York, NY.

Production and Sequencing of the NF-AT120 protein

Protein Preparation

Purification of proteins was carried out at 4°C. Nuclear extracts from Jurkat cells stimulated for 2 hours with PMA at 50 ng/ml and A23187 at 1.0 μM were prepared by the method of Dignam et al. (1983) Nucleic Acids Res., 11:1475-1489, using 0.2 % NP-40 to disrupt the cells. Purification of the 120 kDa component of NF-AT was as follows. Nuclear extracts from PMA/A23187-stimulated Jurkat cells (5 x 10^{10} cells) were fractionated by addition of solid ammonium sulfate (40% saturation) and then centrifuged at 10,000 x g for 20 min. The pellet was dissolved with 7-8 ml of Buffer A (20 mM HEPES (pH 7.9), 0.1 M KCl, 1 mM EDTA, 1 mM EGTA, 5% glycerol, 1 mM DTT, 10 μg/ml leupeptin, and 0.1 mM PMSF), and 2 ml of the sample was applied to a SUPEROSE 6 gel filtration column (HR 16/50, Pharmacia Co.) on an FPLC system at a flow rate of 0.3 ml/min. Each fraction was subjected to mobility shift assay using ^{32}P-labeled NF-AT oligonucleotide as a probe. NP-40 and polydIdC were added to the fractions containing NF-AT binding at 0.2 % and 13 μg/ml, respectively.

The sample, after gel-filtration, was divided into five parts, and each of the protein solutions was applied to NF-AT oligonucleotide-coupled SEPHAROSE 4B (1 ml gel volume) which had been equilibrated with Buffer A containing 0.2 % NP-40 (Buffer B) at gravity flow. The column was washed with more than 10 column volumes of Buffer B containing 1 M urea and then 10 column volumes of Buffer B containing 0.2 M KCl, and was then eluted with Buffer B containing 0.3 M KCl.

Fractions containing NF-AT binding activity were collected and dialyzed against Buffer C (6 M urea, 0.05 M KCl, 20 mM HEPES (pH 7.9), 1 mM EGTA,

1 mM EDTA). The dialysate was applied to a Mono Q column (HR 5/10, Pharmacia Co.) on an FPLC system which had been equilibrated with Buffer C. Elution was by KCl gradient (0.05 M - 0.8 M); 0.5 ml fractions were collected and assayed by mobility shift assay.

Affinity-purified AP-1 was prepared from nuclear extracts of Jurkat cells stimulated with PMA/A23187 by the method of Lee et al. (1987) Cell, 49:741-52, with slight modifications. Recombinant cJun and cFos proteins, which were expressed as histidine fusion proteins in *E. coli* and purified by nickel chelate chromatography (see Abate et al. (1990) Proc. Natl. Acad. Sci., 87:1032-1036), were kindly provided by T. Curran and T. Kerppola (Roche Institute of Molecular Biology).

Oligonucleotides

The oligonucleotides used for competition and mobility shift assays contained the following sequences (only one strand is shown; sequence overhangs are in lower-case letters):

CLE0 element: 5'-gatcGTCACCATTAATCATTTCCTCTAACTGT-3'; see Miyatake et al. (1991) Mol. Cell. Biol., 11:5894-901, and SEQ ID NO: 25;

AP-1 site: 5'-tcgaGCTATGACTCATCCG-3'; see Nakabeppu et al. (1988) Cell, 55:907-15, and SEQ ID NO: 26;

NF-AT site: 5'-gatcGGAGGAAAAACTGTTTCATACAGAAGGCGT-3'; see Emmel et al. (1989) Science, 246:1617-20, and SEQ ID NO: 27.

Mobility shift assay

The DNA binding reactions were performed at room temperature for 30 min in a solution (10 µl) containing 10 mM HEPES (pH 7.9), 1 mM EDTA, 1 mM DTT, 10% glycerol, KCl adjusted to 100 mM, 0.5 ng of ^{32}P-labeled probe (50,000 cpm), and protein sample as indicated in the figure legend, with or without 100 ng/µl of polydIdC. Samples were analyzed on 4% native polyacrylamide gels (Tris-glycine-EDTA buffer) at 120 v, which were dried and exposed to Kodak X-OMAT film.

Denaturation and renaturation of the 120 kDa component of NF-AT

Denaturation and renaturation was as described by McCaffrey et al. (1993) J. Biol. Chem., 268:3747-3752, with slight modifications. The fraction from Mono Q column chromatography which contained the NF-AT120 was concentrated with a microconcentrator (CENTRICON™30, Amicon) and subjected to SDS- 7.5% polyacrylamide gel electrophoresis along with a prestained molecular weight marker. The gel piece corresponding to NF-AT120 was excised and eluted overnight with 200 µl of 20 mM HEPES (pH 7.9), 1 mM EDTA, 1 mM EGTA, 2 mM DTT, 5% glycerol, 0.1% SDS and 0.1 mg/ml BSA. The eluted protein was precipitated with 5 volumes of acetone at -20°C and then washed with cold methanol. The protein precipitate was dissolved with 3 µl of 8 M urea, diluted with buffer B, and left overnight at 4°C to renature. Reconstituted NF-AT binding was assayed by mixing 1 µl of renatured 120 kDa protein with 10 ng of affinity-purified Jurkat AP-1 proteins. Mobility shift conditions were as described above in the presence of 100 ng/µl of polydIdC.

Other Methods

NF-AT oligonucleotide-coupled SEPHAROSE 4B (50 µg of DNA per ml of resin) was prepared by the method of Kadonaga and Tjian (1986) Proc. Natl. Acad. Sci. U.S.A., 83:5889-5893 using CNBr-activated SEPHAROSE 4B (Pharmacia Co.). SDS-polyacrylamide gel electrophoresis was performed according to Laemmli (1970) Nature (London), 227:680-685. Protein concentration was determined by the method of Bradford (1976) Anal. Biochem., 72:248-254, using bovine serum albumin as a standard. All other chemicals were reagent grade or better.

A combination of both phorbol ester (PMA) and calcium ionophore (A23187) elicits maximum production of cytokines that include IL-2 and GM-CSF. Production of these cytokines is inhibited by the immunosuppressive drug CsA. Previous studies have shown that the NF-AT binding site in IL-2 promoter is an essential cis-acting element for PMA/ionomycin-dependent expression and that the nuclear protein which binds to this sequence is a target for CsA action.

Recently, the CLE0 element (position -54 to -40) was shown to be essential for induction of transcription at the GM-CSF promoter in a stimulation-dependent manner. Interestingly, the 3'-half of the CLE0 sequence is identical to that of the NF-AT element (Table 2), which is important for NF-AT binding.

One of the CLE0 binding nuclear factors (NF-CLE0γ) is inducible with PMA/A23187 and sensitive to cycloheximide and CsA, suggesting that this factor is related to NF-AT. Table 3 provides further data on the specificity of binding of NF-AT to the CLE0 binding sequence.

Table 2: Comparison of the NF-AT and CLE0 Elements

	NF-AT Element[a]	CLE0 Element[b]
Regulated Gene	IL-2	GM-CSF
Location[c]	h, -272/-289; m,-275/-292	h,-33/-47; m,-40/-54
Sequence Human	TGAAACAGTTTTTCCTCC (1)	ATTAATCATTTCCTC (2)
Murine	TGAAACAAATTTTCCTCC (3)	ATTAATCATTTCCTC (4)
Inducible nuclear factor bound to the cis-element	NF-AT	NF-CLE0γ
Requirement for induction	Ca ionophore/PMA	Ca ionophore/PMA

[a,b] References:

NF-AT Element: Crabtree, *Science* (1989), 243:355-361, Ullman et al., *Ann. Rev. Immunol.* (1990), 8:421-452, and Jain et al., *Nature* (1992), 356:801-804;

CLE0 Element: Arai et al., *Pharmac. Ther.* (1992), 55:303-318, and Miyatake et al., *Mol. Cell. Biol.* (1991), 11:5894-5901.

[c] The NF-AT and CLE0 elements show the sequence of the non-coding strand. Large dots indicate the identical sequence between NF-AT and CLE0 element. h = human; m = murine.

(1) – (4) Sequences of thisTable 2 correspond to (1) SEQ ID NO: 28; (2) residues 7-21 of SEQ ID NO: 29; (3) SEQ ID NO: 30; and (4) SEQ ID NO: 31.

Table 3: Analysis of NF-CLE0 Binding Sequence:
Sequences of oligonucleotides of the wild type CLE0 Sequence and of a series of double base substitutions (GM40.41 to GM52.53)
[The CLE0 Sequence (with flanking nucleotides) is given in SEQ ID NO: 29, where it consists of residues 7-21.]

Probe	Sequence	γ Binding
CLE0	GTCACC[ATTAATCATTTCCTC]TAACTGT	+
GM40.41	----------GA-------	+
GM42.43	----------AA-------	–
GM44.45	----------GG-------	–
GM46.47	----------CG-------	+
GM48.49	----------GG-------	+
GM50.51	----------CC-------	–
GM52.53	-------GA----------	–
GM54.55	-----AC----------	+
GM43.47	----------C----------	–
GM47	----------C---A----------	+

Notes: Dashes show unchanged bases. The CLE0 sequence is boxed. These oligonucleotide probes were labeled, subjected to electromobility shift assay and checked for NF-CLE0γ binding. The results are shown in the right column: (+) indicates that NF-CLE0γ binding was observed; (–) indicates the absence of NF-CLE0γ binding.

To characterize the relationship between NF-CLE0γ and NF-AT, affinity-purified NF-AT were prepared from nuclear extracts of Jurkat cells stimulated with PMA and A23187. After precipitation with ammonium sulfate and gel filtration with SUPEROSE 6, the fractions containing NF-AT-binding activity were applied to the NF-AT oligonucleotide coupled to a SEPHAROSE 4B column. The column was washed, and elution was carried out with 0.3 M KCl-containing buffer. An elution profile of the affinity purified NF-AT, when compared to the protein concentration of each fraction, yielded a single peak.

The eluted fractions were then subjected to a mobility shift assay using ^{32}P-labeled NF-AT oligonucleotides as a probe, which showed that affinity-purified NF-AT shifts the probe as a single band in either the absence (Figure 1,

lanes 2 and 10) or the presence of polydIdC, and the mobility of the band is the same as that of Jurkat nuclear extracts stimulated with PMA/A23187 (Figure 1, lane 1). Consistent with previous studies which showed that the NF-AT contained AP-1 proteins, the mobility shift band of affinity-purified NF-AT was inhibited by oligonucleotides containing NF-AT and AP-1 binding sequences (Figure 1, lanes 2-4). The NF-AT DNA-binding activity was not inhibited by addition of 10 ng of Sp1 oligonucleotide.

Multiple factors of different sequence specificities are induced to bind the CLE0 element. One of them was identified as a transcription factor AP-1 (see Figure 4B, lane 11), which was induced mainly by PMA alone. Another CLE0-binding protein (NF-CLE0γ), mainly observed in the nuclear extracts from Jurkat cells stimulated with both PMA and A23187 (Figure 1, lane 5), is blocked by CsA treatment. As shown in Figure 1, affinity-purified NF-AT can bind the CLE0 probe strongly and shifts the probe to a position that corresponds to NF-CLE0γ. This binding is a specific interaction between NF-AT120 proteins and CLE0 sequence, because it was blocked with NF-AT and AP-1 oligonucleotides (Figure 1, lanes 8 and 9), but not with the Sp1 sequence.

Conversely, NF-AT DNA binding by this protein complex could also be inhibited by a CLE0 sequence (Figure 1, lane 11). These results were consistent with previous results using nuclear extracts from PMA/A23187-stimulated Jurkat cells and strongly suggested that NF-CLE0γ, which was induced by PMA/A23187, was identical to NF-AT.

NF-AT was purified to near homogeneity to determine whether the NF-CLE0γ shared the same component with NF-AT. Affinity-purified NF-AT was dialyzed against a buffer containing 6 M urea, 20 mM HEPES (pH 7.9), 1 mM EGTA, and 1 mM EDTA to dissociate the AP-1 components from the other component of NF-AT, and was then subjected to Mono-Q chromatography in the presence of 6 M urea. After elution by a KCl gradient (0.05 – 0.8 M), each fraction was assayed by mobility shift in either the absence or the presence of affinity-purified AP-1 proteins from activated Jurkat cells. As shown in Figure 2A and Figure 2B, the Mono Q fractions at a position of 0.2 M KCl contained a protein, which reconstituted NF-AT DNA binding activity with AP-1 proteins (fractions 13 – 19).

Analysis by SDS-PAGE of each fraction (Figure 2C) showed that the appearance of the 120 kDa protein band correlated well with the ability to reconstitute NF-AT DNA binding activity in the presence of AP-1 (Figure 2B). Although another 120 kDa protein was eluted at a position with a slightly higher

concentration of KCl (fractions 20 – 24), that protein failed to reconstitute the NF-AT DNA-binding activity. Approximately 1 to 2 μg of the 120 kDa protein was purified from 100 liters of PMA/A23187-stimulated Jurkat cells as a single polypeptide which was more than 95% homogeneous judging from the protein staining of an SDS-PAGE gel (see Figure 4A).

These observations agree with the range of molecular mass reported previously. The DNA-binding component of NF-AT (NF-AT120) was detected in a cytoplasmic fraction of Jurkat cells by elution and renaturation after SDS-PAGE; it has a molecular mass between 94 and 116 kDa. It was also detected by the same method in nuclear extracts of an activated murine T-cell clone, Ar-5, whose apparent molecular mass was between 90 and 125 kDa (NF-ATp).

To eliminate the possibility of additional components derived from the affinity-purified AP-1 used for the reconstitution, whether the purified 120 kDa protein was able to reconstitute the NF-AT DNA-binding activity with recombinant cJun and cFos was tested. As shown in Figure 3, cJun/cFos heterodimer reconstituted the NF-AT complex in combination with the 120 kDa protein and shifted the probe to the same position as the purified Jurkat AP-1 plus the component (lanes 2 and 3) corresponding to the nuclear form of NF-AT (see Figure 4B, lanes 1 and 3). This heterodimer formed a weak DNA binding complex without the 120 kDa protein, whose mobility was faster than that of NF-AT (lane 4). Although cFos alone did not affect the reconstitution of NF-AT DNA binding (lane 7), cJun homodimer gave a weak reconstitution signal (lane 5). These results showed that the affinity of the Jun/Fos heterodimer to the component was apparently higher than that of the Jun/Jun homodimer, which is consistent with previous results that both the Fos and Jun antibodies affected the NF-AT binding; see Jain et al. (1992) Nature 356:801-804.

In order to determine whether the 120 kDa protein is a component of both NF-AT and NF-CLE0γ, the protein band corresponding to 120 kDa on SDS-PAGE after the Mono Q fraction was excised. The protein was eluted from the gel slice, denatured with 8 M urea and renatured by dilution. The 120 kDa protein, eluted and renatured from the gel, reconstituted both NF-AT and CLE0 DNA-binding activity with AP-1 proteins as well as the protein fraction from Mono Q chromatography (Figure 4B). Faint NF-AT binding in Mono Q fraction, detected without the addition of exogenous AP-1 proteins (Figure 2A and Figure 4B, lane 2), is apparently due to a trace amount of contamination of AP-1 proteins in this fraction. This fraction gave weak CLE0 binding activity which was identified as AP-1 protein-CLE0 DNA complex (Figure 4B, lane 7).

However, the renatured 120 kDa protein from the gel did not contain AP-1 DNA-binding activity (Figure 4B, lane 9). These results suggest that the affinity of the NF-AT component to the NF-AT DNA binding sequence, which appears to be below the detection level in this binding assay, is allosterically induced by AP-1 proteins. A recent report showed that NF-ATp from a murine T cell clone bound the NF-AT sequence even in the absence of exogenous proteins, such as AP-1 proteins; see Jain et al. (1992) Nature 356, 801-804.

Direct binding of 120 kDa protein eluted from the gel slice to the NF-AT sequence was not observed. Further characterization of the component will explain this discrepancy. Interestingly, two bands with different mobility were observed in CLE0γ-binding reconstituted with the 120 kDa protein and AP-1 proteins. This may be due to the heterogeneous forms of AP-1 proteins reconstituting the DNA-binding complex.

Although the binding of both native and reconstituted NF-CLE0γ to the CLE0 sequence seemed to be weaker than that of AP-1 *in vitro* (Figure 1, lane 5, and Figure 4B, lanes 8 and 10), NF-AT binding to the CLE0 sequence (CLE0γ) is propably important in the stimulation dependency of this element. The transfection experiments using plasmids with mutations in the 3'-half of the CLE0 element, which is identical to that in the NF-AT sequence (Table 2), completely abolished PMA/A23187-dependent promoter activity.

Thus, the present results demonstrate that NF-AT binds directly to the CLE0 element on the GM-CSF promoter region and that both NF-CLE0γ and NF-AT share the same nuclear component, NF-AT120, which can reconstitute the DNA-binding *in vitro* with AP-1 proteins, including recombinant cJun/cFos heterodimer. These results strongly suggest that the NF-AT may contribute to coordinate regulation of the expression of both IL-2 and GM-CSF genes in T-cells. The procedure for isolation of the 120 kDa component of NF-AT described here is helpful in elucidating the regulation of cytokine gene expression upon T cell activation.

Generation and purification of proteolytic peptides of the NF-AT120 protein

The prep gel slice containing protein was briefly rinsed with water and acetonitrile to remove excess SDS, smashed into tiny fragments, taken to dryness under vacuum on a SpeedVac (Savant), and then solubilized in 0.2 ml 25 mM Tris buffer (pH 7.5) containing 0.7 µg LysC. The cleavage reaction was carried out at 37°C for 24 hours. The reaction mix was spun at 13,000 rpm and loaded onto a 2.1 x 100 mm AQUAPORE RP-300 reversed phase column, and

peptides eluted with a linear 4-44% acetonitrile gradient (with constant 0.1% TFA). Eluting peptides were monitored at 214 nm and were collected by hand.

Determination of the amino acid sequence of peptides of the NF-AT120

Peptide sequences were determined using an Applied Biosystems 477A Sequencer. Fragments provided peptide sequences of SEQ ID NOs: 1 through 5.

Isolation of a DNA clone encoding NF-AT120 protein.

A 437 bp DNA fragment was amplified from Jurkat cDNA by PCR using degenerate oligonucleotides. The two oligonucleotides of SEQ ID NOs: 32 and 33 were used. After an initial denaturing step of 5 min at 95°C, the amplification was performed using the "step-cycle" program on a PCR THERMOCYCLER machine set to denature at 94°C for 1 min, anneal at 40°C for 1 min, and extend at 72°C for 10 sec, for a total of 40 cycles.

The 437 bp fragment was purified and cloned into the pCR™II plasmid vector obtained from Invitrogen Corp. and sequenced by the chain-termination sequencing method.

The resulting sequence showed that the fragment contained the primers at its ends as expected. The fact that the resulting open reading frame contained the sequence KVVFTEK, which corresponds to the L11 peptide of SEQ ID NO: 1, established that this fragment belonged to cDNA encoding our purified 120 kDa protein. SEQ ID NO: 34 represents the NF-AT120 nucleotide sequence and SEQ ID NO: 35 is the encoded amino acid sequence.

This nucleotide segment can be used to screen for related NF-AT proteins by hybridization. This has led to the discovery that there are three related subfamilies, designated class C, class P, and class X; see SEQ ID NO: 36, 38, and 40 respectively for the encoding DNA sequences and SEQ ID NO: 37, 38, and 41 respectively for the amino acid sequences. The X class has been shown in Northern blots to be highly expressed in the thymus.

The purified protein or defined peptides are useful for generating antibodies by standard methods, as described above. Synthetic peptides or purified protein are presented to an immune system to generate monoclonal or polyclonal antibodies. See, e.g.: Coligan (1991 and Supplements) Current Protocols in Immunology, Wiley/Greene; and Harlow and Lane (1989) Antibodies: A Laboratory Manual, Cold Spring Harbor Press. Alternatively, AP-1

is used as a specific binding reagent, and advantage can be taken of its specificity of binding, much as an antibody would be used. In either case, the binding reagent is either labeled as described above, e.g., for fluorescence or otherwise, or immobilized to a substrate for panning methods.

The binding composition is used for screening of an expression library made from a cell line which expresses an NF-AT120 protein. Standard staining techniques are used to detect or sort intracellular or surface-expressed protein, or surface-expressing transformed cells are screened by panning.

Screening of intracellular expression is performed by various staining or immunofluorescence procedures; see also McMahan et al. (1991) EMBO J., 10:2821-2832. For example, on day 0, precoat two-chamber permanox slides with 1 ml per chamber of fibronectin, 10 ng/ml in PBS, for 30 min at room temperature. Rinse once with PBS. Then plate COS cells at 2-3 x 10^5 cells per chamber in 1.5 ml of growth media. Incubate overnight at 37°C.

On day 1 for each sample, prepare 0.5 ml of a solution of 66 µg/ml DEAE-dextran, 66 µM chloroquine, and 4 µg DNA in serum-free DME. For each set, prepare a positive control, e.g., of a known positive cDNA at 1 and 1/200 dilution, and a negative mock. Rinse cells with serum free DME. Add the DNA solution and incubate 5 hours at 37°C. Remove the medium and add 0.5 ml 10% DMSO in DME for 2.5 min. Remove and wash once with DME. Add 1.5 ml growth medium and incubate overnight.

On day 2, change the medium. On day 3 or 4, fix and stain the cells. Rinse the cells twice with Hank's Buffered Saline Solution (HBSS) and fix in 4% paraformaldehyde/glucose for 5 min. Wash 3X with HBSS. The slides may be stored at -80°C after all liquid is removed. For each chamber, 0.5 ml incubations are performed as follows. Add HBSS/saponin(0.1%) with 32 µl/ml of 1M NaN_3 for 20 min. Cells are then washed with HBSS/saponin 1X. Add specific binding reagent, e.g., antibody, to cells and incubate for 30 min. Wash cells twice with HBSS/saponin. Add second antibody, e.g., anti-mouse antibody (e.g., from Vector), at 1/200 dilution, and incubate for 30 min. Prepare ELISA solution, e.g., VECTOR ELITE ABC horseradish peroxidase solution, and preincubate for 30 min. Use, e.g., 1 drop of solution A (avidin) and 1 drop solution B (biotin) per 2.5 ml HBSS/saponin. Wash cells twice with HBSS/saponin. Add ABC HRP solution and incubate for 30 min. Wash cells once with HBSS, and then a second time for 2 min, to close the cells. Then add VECTOR diaminobenzoic acid (DAB) for 5 to 10 min. Use 2 drops of buffer plus 4 drops DAB plus 2 drops of H_2O_2 per 5 ml of glass-distilled water. Carefully remove chamber and rinse

slide in water. Air dry for a few minutes, and then add 1 drop of CRYSTAL MOUNT and a cover slip. Bake for 5 min at 85-90°C.

Alternatively, the binding compositions are used to affinity purify or sort out cells expressing the protein; see, e.g., Sambrook et al. or Ausubel et al.

In another method, the peptide segments are used to predict appropriate oligonucleotides to screen a library. The genetic code is used to select appropriate oligonucleotides useful as probes for directly screening a library; see, e.g., SEQ ID NO: 1 through 24. Alternatively, polymerase chain reaction (PCR) techniques will be applied. Synthetic oligonucleotides in appropriate orientations are used as primers to select correct clones from a library. Various combinations of upstream/downstream sense/antisense combinations are tested until an appropriate clone is amplified and detected. 3'- or 5'-anchor PCR techniques can also be applied.

Another strategy is to screen for a membrane-bound expression product by panning. The cDNA library is constructed in an expression vector which attaches the product to the cell membrane. The soluble binding partner or antibodies raised against the defined peptide fragments can be immobilized and used to immobilize expressing cells. Immobilization may be achieved by use of appropriate antibodies which recognize, e.g., the soluble receptor construct, or by use of antibodies raised against the first antibodies. Recursive cycles of selection and amplification lead to enrichment of appropriate clones and eventual isolation of appropriate expressing clones.

Phage expression libraries can be screened by soluble AP-1 or anti-fragment antibodies. Appropriate label techniques, e.g., antibodies, will allow specific labeling of appropriate clones.

Screening by hybridization using degenerate probes based upon the peptide sequences will also allow isolation of appropriate clones. Alternatively, use of appropriate primers for PCR screening will yield enrichment of appropriate nucleic acid clones.

Similar methods are applicable to isolate either species or allelic variants. Species variants are isolated using cross-species hybridization techniques based upon isolation of a full-length isolate or fragment from one species as a probe. Alternatively, similar assays can be developed, e.g., in mouse cells for isolation of a corresponding NF-AT120 protein.

All references cited herein are incorporated herein by reference to the same extent as if each individual publication or patent application was

specifically and individually indicated to be incorporated by reference in its entirety for all purposes.

Many modifications and variations of this invention can be made without departing from its spirit and scope, as will be apparent to those skilled in the art. The specific embodiments described herein are offered by way of example only, and the invention is to be limited only by the terms of the appended claims, along with the full scope of equivalents to which such claims are entitled.

SEQUENCE LISTING

SEQ ID NO: 1 is NF-AT120 peptide 1 sequence.
SEQ ID NO: 2 is NF-AT120 peptide 2 sequence.
SEQ ID NO: 3 is NF-AT120 peptide 3 sequence.
SEQ ID NO: 4 is NF-AT120 peptide 4 sequence.
SEQ ID NO: 5 is NF-AT120 peptide 5 sequence.
SEQ ID NO: 6 is nucleotide sequence for NF-AT peptide 1.
SEQ ID NO: 7 is nucleotide sequence for NF-AT peptide 2.
SEQ ID NO: 8 is nucleotide sequence for NF-AT peptide 2.
SEQ ID NO: 9 is nucleotide sequence for NF-AT peptide 3.
SEQ ID NO: 10 is nucleotide sequence for NF-AT peptide 3.
SEQ ID NO: 11 is nucleotide sequence for NF-AT peptide 3.
SEQ ID NO: 12 is nucleotide sequence for NF-AT peptide 3.
SEQ ID NO: 13 is nucleotide sequence for NF-AT peptide 4.
SEQ ID NO: 14 is nucleotide sequence for NF-AT peptide 4.
SEQ ID NO: 15 is nucleotide sequence for NF-AT peptide 4.
SEQ ID NO: 16 is nucleotide sequence for NF-AT peptide 4.
SEQ ID NO: 17 is nucleotide sequence for NF-AT peptide 5.
SEQ ID NO: 18 is nucleotide sequence for NF-AT peptide 5.
SEQ ID NO: 19 is nucleotide sequence for NF-AT peptide 5.
SEQ ID NO: 20 is nucleotide sequence for NF-AT peptide 5.
SEQ ID NO: 21 is nucleotide sequence for NF-AT peptide 5.
SEQ ID NO: 22 is nucleotide sequence for NF-AT peptide 5.
SEQ ID NO: 23 is nucleotide sequence for NF-AT peptide 5.
SEQ ID NO: 24 is nucleotide sequence for NF-AT peptide 5.
SEQ ID NO: 25 is CLE0 nucleotide sequence.
SEQ ID NO: 26 is AP-1 site.
SEQ ID NO: 27 is NF-AT site.
SEQ ID NO: 28 is huNF-AT element.
SEQ ID NO: 29 is huCLE0 element.
SEQ ID NO: 30 is moNF-AT element.
SEQ ID NO: 31 is moCLE0 element.
SEQ ID NO: 32 is L27-1S oligonucleotide.
SEQ ID NO: 33 is L33-4A oligonucleotide.
SEQ ID NO: 34 is huNF-AT120, P subfamily, nucleotide sequence
SEQ ID NO: 35 is huNF-AT120, P subfamily, amino acid sequence
SEQ ID NO: 36 is huNF-AT120, C subfamily, nucleotide sequence
SEQ ID NO: 37 is huNF-AT120, C subfamily, amino acid sequence
SEQ ID NO: 38 is huNF-AT120, P subfamily, nucleotide sequence
SEQ ID NO: 39 is huNF-AT120, P subfamily, amino acid sequence
SEQ ID NO: 40 is huNF-AT120, X subfamily, nucleotide sequence
SEQ ID NO: 41 is huNF-AT120, X subfamily, amino acid sequence

(1) GENERAL INFORMATION:

(i) INVENTOR: Arai, Naoko
Masuda, Esteban S.
Tokumitsu, Hiroshi

(i) APPLICANT: Schering Corporation

(ii) TITLE OF INVENTION: PURIFIED COMPONENTS OF MAMMALIAN TRANSCRIPTION REGULATION COMPLEXES, AND ANALOGS

(iii) NUMBER OF SEQUENCES: 41

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: John H. C. Blasdale, Schering-Plough Corporation, M-3-W
(B) STREET: One Giralda Farms
(C) CITY: Madison
(D) STATE: New Jersey
(E) COUNTRY: USA
(F) ZIP: 07940-1000

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: Apple Macintosh IIci
(C) OPERATING SYSTEM: System Software 7.1
(D) SOFTWARE: Microsoft Word 5.1a

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: PCT/US 94/
(B) FILING DATE:
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/222,626
(B) FILING DATE: 04-APR-1994

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/148,061
(B) FILING DATE: 05-NOV-1993

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/113,971
(B) FILING DATE: 30-AUG-1993

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/099,998
(B) FILING DATE: 30-JUL-1993

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/088,483
(B) FILING DATE: 06-JUL-1993

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Blasdale, John H. C.
(B) REGISTRATION NUMBER: 31,895
(C) REFERENCE/DOCKET NUMBER: DX0392K4

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 201-822-7398
(B) TELEFAX: 201-822-7039

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Lys Val Val Phe Thr Glu Lys
1 5

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Ala Arg Asn Gln Thr Pro Ser Lys
1 5

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Val Leu Glu Ile Pro Leu Glu Pro Lys
1 5

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 1..2
(D) OTHER INFORMATION: /label= variation
/note= "can also be Ser Gln"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Trp Gln Pro Asn Met Leu Phe Val Glu Ile Pro Glu Tyr Arg Asn
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Gln Glu Gln Asn Leu Asp Gln Thr Tyr Leu Asp Gln Val Asn Glu Ile
1 5 10 15

Val Arg Lys

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

AARGTNGTNT TYACNGARAA 20

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GCNAGRAAYC ARACNCC 17

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GCNCGNAAYC ARACNCC 17

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GTNTTRGARA THCCNTTRGA RCCNAAR 27

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GTNTTRGARA THCCNCUNGA RCCNAAR 27

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GTNCUNGARA THCCNTTRGA RCCNAAR 27

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GTNCUNGARA THCCNCUNGA RCCNAAR 27

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 45 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

TGGCARCCNA AYAUGTTRTT YGTNGARATM CCNGARTAYC GNAAY 45

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 45 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

TGGCARCCNA AYAUGTTRTT YGTNGARATM CCNGARTAYA GRAAY 45

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 45 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

TGGCARCCNA AYAUGCUNTT YGTNGARATM CCNGARTAYC GNAAY 45

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 45 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

TGGCARCCNA AYAUGCUNTT YGTNGARATM CCNGARTAYA GRAAY 45

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 57 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

CARGARCARA AYTTRGAYCA RACNTAYTTR CAYCARGTNA AYGARATHGT NCGNAAR 57

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 57 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CARGARCARA AYTTRGAYCA RACNTAYTTR CAYCARGTNA AYGARATHGT NAGRAAR 57

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 57 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

CARGARCARA AYTTRGAYCA RACNTAYCTN CAYCARGTNA AYGARATHGT NCGNAAR 57

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 57 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

CARGARCARA AYTTRGAYCA RACNTAYCTN CAYCARGTNA AYGARATHGT NAGRAAR 57

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 57 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

CARGARCARA AYCTNGAYCA RACNTAYTTR CAYCARGTNA AYGARATHGT NCGNAAR 57

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 57 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

CARGARCARA AYCTNGAYCA RACNTAYTTR CAYCARGTNA AYGARATHGT NAGRAAR 57

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 57 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

CARGARCARA AYCTNGAYCA RACNTAYCTN CAYCARGTNA AYGARATHGT NCGNAAR 57

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 57 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

CARGARCARA AYCTNGAYCA RACNTAYCTN CAYCARGTNA AYGARATHGT NAGRAAR 57

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

GATCGTCACC ATTAATCATT TCCTCTAACT GT 32

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

TCGAGCTATG ACTCATCCG 19

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 34 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GATCGGAGGA AAAACTGTTT CATACAGAAG GCGT 34

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

TGAAACAGTT TTTCCTCC 18

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(20..21, "ga")
(D) OTHER INFORMATION: /standard_name= "GM40.41"

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(18..19, "aa")
(D) OTHER INFORMATION: /standard_name= "GM42.43"

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(16..17, "gg")
(D) OTHER INFORMATION: /standard_name= "GM44.45"

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(14..15, "cg")
(D) OTHER INFORMATION: /standard_name= "GM46.47"

(ix) FEATURE:
(A) NAME/KEY: misc_difference

(B) LOCATION: replace(12..13, "gg")
(D) OTHER INFORMATION: /standard_name= "GM48.49"

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(10..11, "cc")
(D) OTHER INFORMATION: /standard_name= "GM50.51"

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(8..9, "ga")
(D) OTHER INFORMATION: /standard_name= "GM52.53"

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(6..7, "ac")
(D) OTHER INFORMATION: /standard_name= "GM54.55"

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(14..18, "cttta")
(D) OTHER INFORMATION: /standard_name= "GM43.47"

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(14..15, "ct")
(D) OTHER INFORMATION: /standard_name= "GM47"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

GTCACCATTA ATCATTTCCT CTAACTGT 28

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

TGAAACAAAT TTTCCTCC 18

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

ATTAATCATT TCCTC 15

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

GARATHCCNY TNGARCC 17

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

TAYTCNGGDA TYTCNACRAA 20

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 420 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..420

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

```
GAG ATT CCG TTG GAG CCC AAA AAC AAC ATG AGG GCA ACC ATC GAC TGT      48
Glu Ile Pro Leu Glu Pro Lys Asn Asn Met Arg Ala Thr Ile Asp Cys
 1               5                  10                  15

GCG GGG ATC TTG AAG CTT AGA AAC GCC GAT ATT GAG CTG CGG AAA GGC      96
Ala Gly Ile Leu Lys Leu Arg Asn Ala Asp Ile Glu Leu Arg Lys Gly
            20                  25                  30

GAG ACG GAC ATT GGA AGA AAG AAC ACG CGG GTG AGA CTG GTT TTC CGA     144
Glu Thr Asp Ile Gly Arg Lys Asn Thr Arg Val Arg Leu Val Phe Arg
        35                  40                  45

GTT CAC ATC CCA GAG TCC AGT GGC AGA ATC GTC TCT TTA CAG ACT GCA     192
Val His Ile Pro Glu Ser Ser Gly Arg Ile Val Ser Leu Gln Thr Ala
    50                  55                  60

TCT AAC CCC ATC GAG TGC TCC CAG CGA TCT CGT CAC GAG CTG CCC ATG     240
Ser Asn Pro Ile Glu Cys Ser Gln Arg Ser Arg His Glu Leu Pro Met
65                  70                  75                  80

GTT GAA AGA CAA GAC ACA GAC AGC TGC CTG GTC TAT GGC GGC CAG CAA     288
Val Glu Arg Gln Asp Thr Asp Ser Cys Leu Val Tyr Gly Gly Gln Gln
                85                  90                  95

ATG ATC CTC ACG GGG CAG AAC TTT ACA TCC GAG TCC AAA GTT GTG TTT     336
Met Ile Leu Thr Gly Gln Asn Phe Thr Ser Glu Ser Lys Val Val Phe
            100                 105                 110

ACT GAG AAG ACC ACA GAT GGA CAG CAA ATT TGG GAG ATG GAA GCC ACG     384
Thr Glu Lys Thr Thr Asp Gly Gln Gln Ile Trp Glu Met Glu Ala Thr
        115                 120                 125

GTG GAT AAG GAC AAG AGC CAG CCC AAC ATG CTT TTC                     420
Val Asp Lys Asp Lys Ser Gln Pro Asn Met Leu Phe
    130                 135                 140
```

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 140 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

```
Glu Ile Pro Leu Glu Pro Lys Asn Asn Met Arg Ala Thr Ile Asp Cys
 1               5                  10                  15

Ala Gly Ile Leu Lys Leu Arg Asn Ala Asp Ile Glu Leu Arg Lys Gly
            20                  25                  30
```

Glu Thr Asp Ile Gly Arg Lys Asn Thr Arg Val Arg Leu Val Phe Arg
35 40 45

Val His Ile Pro Glu Ser Ser Gly Arg Ile Val Ser Leu Gln Thr Ala
50 55 60

Ser Asn Pro Ile Glu Cys Ser Gln Arg Ser Arg His Glu Leu Pro Met
65 70 75 80

Val Glu Arg Gln Asp Thr Asp Ser Cys Leu Val Tyr Gly Gly Gln Gln
85 90 95

Met Ile Leu Thr Gly Gln Asn Phe Thr Ser Glu Ser Lys Val Val Phe
100 105 110

Thr Glu Lys Thr Thr Asp Gly Gln Gln Ile Trp Glu Met Glu Ala Thr
115 120 125

Val Asp Lys Asp Lys Ser Gln Pro Asn Met Leu Phe
130 135 140

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2853 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 340..2490

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

CGGACGCGTG GGCTTTAAAA AGGCAGGAGG CAGAGCGCGG CCCTGCGTCA GAGCGAGACT 60

CAGAGGCTCC GAACTCGCCG GCGGAGTCGC CGCGCCAGAT CCCAGCAGCA GGGCGCGGGC 120

ACCGGGGCGC GGGCAGGGCT CGGAGCCACC GCGCAGGTCC TAGGGCCGCG GCCGGGCCCC 180

GCCACGCGCG CACACGCCCC TCGATGACTT TCCTCCGGGG CGCGCGGCGC TGAGCCCGGG 240

GCGAGGGCTG TCTTCCCGGA GACCCGACCC CGGCAGCGCG GGGCGGCCAC TTCTCCTGTG 300

CCTCCGCCCG CTGCTCCACT CCCCGCCGCC GCCGCGCGG ATG CCA AGC ACC AGC 354
Met Pro Ser Thr Ser
1 5

TTT CCA GTC CCT TCC AAG TTT CCA CTT GGC CCT GCG GCT GCG GTC TTC 402
Phe Pro Val Pro Ser Lys Phe Pro Leu Gly Pro Ala Ala Ala Val Phe
10 15 20

```
GGG AGA GGA GAA ACT TTG GGG CCC GCG CCG CGC GCC GGC GGC ACC ATG      450
Gly Arg Gly Glu Thr Leu Gly Pro Ala Pro Arg Ala Gly Gly Thr Met
            25                  30                  35

AAG TCA GCG GAG GAA GAA CAC TAT GGC TAT GCA TCC TCC AAC GTC AGC      498
Lys Ser Ala Glu Glu Glu His Tyr Gly Tyr Ala Ser Ser Asn Val Ser
        40                  45                  50

CCC GCC CTG CCG CTC CCC ACG GCG CAC TCC ACC CTG CCG GCC CCG TGC      546
Pro Ala Leu Pro Leu Pro Thr Ala His Ser Thr Leu Pro Ala Pro Cys
    55                  60                  65

CAC AAC CTT CAG ACC TCC ACA CCG GGC ATC ATC CCG CCG GCG GAT CAC      594
His Asn Leu Gln Thr Ser Thr Pro Gly Ile Ile Pro Pro Ala Asp His
70                  75                  80                  85

CCC TCG GGG TAC GGA GCA GCT TTG GAC GGT GGG CCC GCG GGC TAC TTC      642
Pro Ser Gly Tyr Gly Ala Ala Leu Asp Gly Gly Pro Ala Gly Tyr Phe
                90                  95                  100

CTC TCC TCC GGC CAC ACC AGG CCT GAT GGG GCC CCT GCC CTG GAG AGT      690
Leu Ser Ser Gly His Thr Arg Pro Asp Gly Ala Pro Ala Leu Glu Ser
            105                 110                 115

CCT CGC ATC GAG ATA ACC TCG TGC TTG GGC CTG TAC CAC AAC AAT AAC      738
Pro Arg Ile Glu Ile Thr Ser Cys Leu Gly Leu Tyr His Asn Asn Asn
        120                 125                 130

CAG TTT TTC CAC GAT GTG GAG GTG GAA GAC GTC CTC CCT AGC TCC AAA      786
Gln Phe Phe His Asp Val Glu Val Glu Asp Val Leu Pro Ser Ser Lys
    135                 140                 145

CGG TCC CCC TCC ACG GCC ACG CTG AGT CTG CCC AGC CTG GAG GCC TAC      834
Arg Ser Pro Ser Thr Ala Thr Leu Ser Leu Pro Ser Leu Glu Ala Tyr
150                 155                 160                 165

AGA GAC CCC TCG TGC CTG AGC CCG GCC AGC AGC CTG TCC TCC CGG AGC      882
Arg Asp Pro Ser Cys Leu Ser Pro Ala Ser Ser Leu Ser Ser Arg Ser
                170                 175                 180

TGC AAC TCA GAG GCC TCC TCC TAC GAG TCC AAC TAC TCG TAC CCG TAC      930
Cys Asn Ser Glu Ala Ser Ser Tyr Glu Ser Asn Tyr Ser Tyr Pro Tyr
            185                 190                 195

GCG TCC CCC CAG ACG TCG CCA TGG CAG TCT CCC TGC GTG TCT CCC AAG      978
Ala Ser Pro Gln Thr Ser Pro Trp Gln Ser Pro Cys Val Ser Pro Lys
        200                 205                 210

ACC ACG GAC CCC GAG GAG GGC TTT CCC CGC GGG CTG GGG GCC TGC ACA     1026
Thr Thr Asp Pro Glu Glu Gly Phe Pro Arg Gly Leu Gly Ala Cys Thr
    215                 220                 225

CTG CTG GGT TCC CCG CAG CAC TCC CCC TCC ACC TCG CCC CGC GCC AGC     1074
Leu Leu Gly Ser Pro Gln His Ser Pro Ser Thr Ser Pro Arg Ala Ser
230                 235                 240                 245
```

```
GTC ACT GAG GAG AGC TGG CTG GGT GCC CGC TCC TCC AGA CCC GCG TCC      1122
Val Thr Glu Glu Ser Trp Leu Gly Ala Arg Ser Ser Arg Pro Ala Ser
                250                 255                 260

CCT TGC AAC AAG AGG AAG TAC AGA CTC AAC GGC CGG CAG CCG CCC TAC      1170
Pro Cys Asn Lys Arg Lys Tyr Arg Leu Asn Gly Arg Gln Pro Pro Tyr
            265                 270                 275

TCA CCC CAC CAC TCG CCC ACG CCG TCC CCG CAC GGC TCC CCG CGG GTC      1218
Ser Pro His His Ser Pro Thr Pro Ser Pro His Gly Ser Pro Arg Val
        280                 285                 290

AGC GTG ACC GAC GAC TCG TGG TTG GGC AAC ACC ACC CAG TAC ACC AGC      1266
Ser Val Thr Asp Asp Ser Trp Leu Gly Asn Thr Thr Gln Tyr Thr Ser
    295                 300                 305

TCG GCC ATC GTG GCC GCC ATC AAC GCG CTG ACC ACC GAC AGC AGC CTG      1314
Ser Ala Ile Val Ala Ala Ile Asn Ala Leu Thr Thr Asp Ser Ser Leu
310                 315                 320                 325

GAC CTG GGA GAT GGC GTC CCT GTC AAG TCC CGC AAG ACC ACC CTG GAG      1362
Asp Leu Gly Asp Gly Val Pro Val Lys Ser Arg Lys Thr Thr Leu Glu
                330                 335                 340

CAG CCG CCC TCA GTG GCG CTC AAG GTG GAG CCC GTC GGG GAG GAC CTG      1410
Gln Pro Pro Ser Val Ala Leu Lys Val Glu Pro Val Gly Glu Asp Leu
            345                 350                 355

GGC AGC CCC CCG CCC CCG GCC GAC TTC GCG CCC GAA GAC TAC TCC TCT      1458
Gly Ser Pro Pro Pro Pro Ala Asp Phe Ala Pro Glu Asp Tyr Ser Ser
        360                 365                 370

TTC CAG CAC ATC AGG AAG GGC GGC TTC TGC GAC CAG TAC CTG GCG GTG      1506
Phe Gln His Ile Arg Lys Gly Gly Phe Cys Asp Gln Tyr Leu Ala Val
    375                 380                 385

CCG CAG CAC CCC TAC CAG TGG GCG AAG CCC AAG CCC CTG TCC CCT ACG      1554
Pro Gln His Pro Tyr Gln Trp Ala Lys Pro Lys Pro Leu Ser Pro Thr
390                 395                 400                 405

TCC TAC ATG AGC CCG ACC CTG CCC GCC CTG GAC TGG CAG CTG CCG TCC      1602
Ser Tyr Met Ser Pro Thr Leu Pro Ala Leu Asp Trp Gln Leu Pro Ser
                410                 415                 420

CAC TCA GGC CCG TAT GAG CTT CGG ATT GAG GTG CAG CCC AAG TCC CAC      1650
His Ser Gly Pro Tyr Glu Leu Arg Ile Glu Val Gln Pro Lys Ser His
            425                 430                 435

CAC CGA GCC CAC TAC GAG ACG GAG GGC AGC CGG GGG GCC GTG AAG GCG      1698
His Arg Ala His Tyr Glu Thr Glu Gly Ser Arg Gly Ala Val Lys Ala
        440                 445                 450

TCG GCC GGA GGA CAC CCC ATC GTG CAG CTG CAT GGC TAC TTG GAG AAT      1746
Ser Ala Gly Gly His Pro Ile Val Gln Leu His Gly Tyr Leu Glu Asn
    455                 460                 465
```

```
GAG CCG CTG ATG CTG CAG CTT TTC ATT GGG ACG GCG GAC GAC CGC CTG      1794
Glu Pro Leu Met Leu Gln Leu Phe Ile Gly Thr Ala Asp Asp Arg Leu
470                 475                 480                 485

CTG CGC CCG CAC GCC TTC TAC CAG GTG CAC CGC ATC ACA GGG AAG ACC      1842
Leu Arg Pro His Ala Phe Tyr Gln Val His Arg Ile Thr Gly Lys Thr
                490                 495                 500

GTG TCC ACC ACC AGC CAC GAG GCT ATC CTC TCC AAC ACC AAA GTC CTG      1890
Val Ser Thr Thr Ser His Glu Ala Ile Leu Ser Asn Thr Lys Val Leu
            505                 510                 515

GAG ATC CCA CTC CTG CCG GAG AAC AGC ATG CGA GCC GTC ATT GAC TGT      1938
Glu Ile Pro Leu Leu Pro Glu Asn Ser Met Arg Ala Val Ile Asp Cys
        520                 525                 530

GCC GGA ATC CTG AAA CTC AGA AAC TCC GAC ATT GAA CTT CGG AAA GGA      1986
Ala Gly Ile Leu Lys Leu Arg Asn Ser Asp Ile Glu Leu Arg Lys Gly
    535                 540                 545

GAG ACG GAC ATC GGG AGG AAG AAC ACA CGG GTA CGG CTG GTG TTC CGC      2034
Glu Thr Asp Ile Gly Arg Lys Asn Thr Arg Val Arg Leu Val Phe Arg
550                 555                 560                 565

GTT CAC GTC CCG CAA CCC AGC GGC CGC ACG CTG TCC CTG CAG GTG GCC      2082
Val His Val Pro Gln Pro Ser Gly Arg Thr Leu Ser Leu Gln Val Ala
                570                 575                 580

TCC AAC CCC ATC GAA TGC TCC CAG CGC TCA GCT CAG GAG CTG CCT CTG      2130
Ser Asn Pro Ile Glu Cys Ser Gln Arg Ser Ala Gln Glu Leu Pro Leu
            585                 590                 595

GTG GAG AAG CAG AGC ACG GAC AGC TAT CCG GTC GTG GGC GGG AAG AAG      2178
Val Glu Lys Gln Ser Thr Asp Ser Tyr Pro Val Val Gly Gly Lys Lys
        600                 605                 610

ATG GTC CTG TCT GGC CAC AAC TTC CTG CAG GAC TCC AAG GTC ATT TTC      2226
Met Val Leu Ser Gly His Asn Phe Leu Gln Asp Ser Lys Val Ile Phe
    615                 620                 625

GTG GAG AAA GCC CCA GAT GGC CAC CAT GTC TGG GAG ATG GAA GCG AAA      2274
Val Glu Lys Ala Pro Asp Gly His His Val Trp Glu Met Glu Ala Lys
630                 635                 640                 645

ACT GAC CGG GAC CTG TGC AAG CCG AAT TCT CTG GTG GTT GAG ATC CCG      2322
Thr Asp Arg Asp Leu Cys Lys Pro Asn Ser Leu Val Val Glu Ile Pro
                650                 655                 660

CCA TTT CGG AAT CAG AGG ATA ACC AGC CCC GTT CAC GTC AGT TTC TAC      2370
Pro Phe Arg Asn Gln Arg Ile Thr Ser Pro Val His Val Ser Phe Tyr
            665                 670                 675

GTC TGC AAC GGG AAG AGA AAG CGA AGC CAG TAC CAG CGT TTC ACC TAC      2418
Val Cys Asn Gly Lys Arg Lys Arg Ser Gln Tyr Gln Arg Phe Thr Tyr
        680                 685                 690
```

CTT CCC GCC AAC GGT AAC GCC ATC TTT CTA ACC GTA AGC CGT GAA CAT 2466
Leu Pro Ala Asn Gly Asn Ala Ile Phe Leu Thr Val Ser Arg Glu His
695 700 705

GAG CGC GTG GGG TGC TTT TTC TAAAGACGCA GAAACGACGT CGCCGTAAAG 2517
Glu Arg Val Gly Cys Phe Phe
710 715

CAGCGTGGCG TGTTGCACAT TTAACTGTGT GATGTCCCGT TAGTGAGACC GAGCCATCGA 2577

TGCCCTGAAA AGGAAAGGAA AAGGGAAGCT TCGGATGCAT TTTCCTTGAT CCCTGTTGGG 2637

GGTGGGGGGC GGGGGTTGCA TACTCAGATA GTCACGGTTA TTTTGCTTCT TGCGAATGTA 2697

TAACAGCCAA GGGGAAAACA TGGCTCTTCT GCTCCAAAAA ACTGAGGGGG TCCTGGTGTG 2757

CATTTGCACC CTAAAGCTGC TTACGGTGAA AAGGCAAATA GGTATAGCTA TTTTGCAGGC 2817

ACCTTTAGGA ATAAACTTTG CTTTTAAAAA AAAAAA 2853

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 716 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

Met Pro Ser Thr Ser Phe Pro Val Pro Ser Lys Phe Pro Leu Gly Pro
1 5 10 15

Ala Ala Ala Val Phe Gly Arg Gly Glu Thr Leu Gly Pro Ala Pro Arg
20 25 30

Ala Gly Gly Thr Met Lys Ser Ala Glu Glu Glu His Tyr Gly Tyr Ala
35 40 45

Ser Ser Asn Val Ser Pro Ala Leu Pro Leu Pro Thr Ala His Ser Thr
50 55 60

Leu Pro Ala Pro Cys His Asn Leu Gln Thr Ser Thr Pro Gly Ile Ile
65 70 75 80

Pro Pro Ala Asp His Pro Ser Gly Tyr Gly Ala Ala Leu Asp Gly Gly
85 90 95

Pro Ala Gly Tyr Phe Leu Ser Ser Gly His Thr Arg Pro Asp Gly Ala
100 105 110

Pro Ala Leu Glu Ser Pro Arg Ile Glu Ile Thr Ser Cys Leu Gly Leu
115 120 125

Tyr His Asn Asn Asn Gln Phe Phe His Asp Val Glu Val Glu Asp Val
130 135 140

Leu Pro Ser Ser Lys Arg Ser Pro Ser Thr Ala Thr Leu Ser Leu Pro
145 150 155 160

Ser Leu Glu Ala Tyr Arg Asp Pro Ser Cys Leu Ser Pro Ala Ser Ser
165 170 175

Leu Ser Ser Arg Ser Cys Asn Ser Glu Ala Ser Ser Tyr Glu Ser Asn
180 185 190

Tyr Ser Tyr Pro Tyr Ala Ser Pro Gln Thr Ser Pro Trp Gln Ser Pro
195 200 205

Cys Val Ser Pro Lys Thr Thr Asp Pro Glu Glu Gly Phe Pro Arg Gly
210 215 220

Leu Gly Ala Cys Thr Leu Leu Gly Ser Pro Gln His Ser Pro Ser Thr
225 230 235 240

Ser Pro Arg Ala Ser Val Thr Glu Glu Ser Trp Leu Gly Ala Arg Ser
245 250 255

Ser Arg Pro Ala Ser Pro Cys Asn Lys Arg Lys Tyr Arg Leu Asn Gly
260 265 270

Arg Gln Pro Pro Tyr Ser Pro His His Ser Pro Thr Pro Ser Pro His
275 280 285

Gly Ser Pro Arg Val Ser Val Thr Asp Asp Ser Trp Leu Gly Asn Thr
290 295 300

Thr Gln Tyr Thr Ser Ser Ala Ile Val Ala Ala Ile Asn Ala Leu Thr
305 310 315 320

Thr Asp Ser Ser Leu Asp Leu Gly Asp Gly Val Pro Val Lys Ser Arg
325 330 335

Lys Thr Thr Leu Glu Gln Pro Pro Ser Val Ala Leu Lys Val Glu Pro
340 345 350

Val Gly Glu Asp Leu Gly Ser Pro Pro Pro Pro Ala Asp Phe Ala Pro
355 360 365

Glu Asp Tyr Ser Ser Phe Gln His Ile Arg Lys Gly Gly Phe Cys Asp
370 375 380

Gln Tyr Leu Ala Val Pro Gln His Pro Tyr Gln Trp Ala Lys Pro Lys
385 390 395 400

Pro Leu Ser Pro Thr Ser Tyr Met Ser Pro Thr Leu Pro Ala Leu Asp
405 410 415

Trp Gln Leu Pro Ser His Ser Gly Pro Tyr Glu Leu Arg Ile Glu Val
420 425 430

Gln Pro Lys Ser His His Arg Ala His Tyr Glu Thr Glu Gly Ser Arg
435 440 445

Gly Ala Val Lys Ala Ser Ala Gly Gly His Pro Ile Val Gln Leu His
450 455 460

Gly Tyr Leu Glu Asn Glu Pro Leu Met Leu Gln Leu Phe Ile Gly Thr
465 470 475 480

Ala Asp Asp Arg Leu Leu Arg Pro His Ala Phe Tyr Gln Val His Arg
485 490 495

Ile Thr Gly Lys Thr Val Ser Thr Thr Ser His Glu Ala Ile Leu Ser
500 505 510

Asn Thr Lys Val Leu Glu Ile Pro Leu Leu Pro Glu Asn Ser Met Arg
515 520 525

Ala Val Ile Asp Cys Ala Gly Ile Leu Lys Leu Arg Asn Ser Asp Ile
530 535 540

Glu Leu Arg Lys Gly Glu Thr Asp Ile Gly Arg Lys Asn Thr Arg Val
545 550 555 560

Arg Leu Val Phe Arg Val His Val Pro Gln Pro Ser Gly Arg Thr Leu
565 570 575

Ser Leu Gln Val Ala Ser Asn Pro Ile Glu Cys Ser Gln Arg Ser Ala
580 585 590

Gln Glu Leu Pro Leu Val Glu Lys Gln Ser Thr Asp Ser Tyr Pro Val
595 600 605

Val Gly Gly Lys Lys Met Val Leu Ser Gly His Asn Phe Leu Gln Asp
610 615 620

Ser Lys Val Ile Phe Val Glu Lys Ala Pro Asp Gly His His Val Trp
625 630 635 640

Glu Met Glu Ala Lys Thr Asp Arg Asp Leu Cys Lys Pro Asn Ser Leu
645 650 655

Val Val Glu Ile Pro Pro Phe Arg Asn Gln Arg Ile Thr Ser Pro Val
660 665 670

His Val Ser Phe Tyr Val Cys Asn Gly Lys Arg Lys Arg Ser Gln Tyr
675 680 685

Gln Arg Phe Thr Tyr Leu Pro Ala Asn Gly Asn Ala Ile Phe Leu Thr
690 695 700

Val Ser Arg Glu His Glu Arg Val Gly Cys Phe Phe
705 710 715

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2839 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 30..2129

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: replace(1574..1576, "gtt")

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: replace(1574..1576, "gat")

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: replace(1574..1576, "ggt")

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: replace(1833..1835, "ctc")

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: replace(1833..1835, "ccc")

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: replace(1833..1835, "cac")

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: replace(1864..1866, "tta")

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: replace(1869..1871, "tcc")

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: replace(1869..1871, "tac")

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: replace(1869..1871, "tgc")

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: replace(2824..2826, "atg")

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

```
CTCATTATTC CCCCAGAACC TCGCCAATA ATG TCA CCT CGA ACC AGC CTC GCC        53
                                Met Ser Pro Arg Thr Ser Leu Ala
                                  1               5

GAT GAC AGC TGC CTG GGC CGC CAC TCG CCC GTG CCC CGT CCG GCC TCC      101
Asp Asp Ser Cys Leu Gly Arg His Ser Pro Val Pro Arg Pro Ala Ser
     10                  15                  20

CGC TCC TCA TCG CCT GGT GCC AAG CGG AGG CAT TCG TGC GCC GAG GCC      149
Arg Ser Ser Ser Pro Gly Ala Lys Arg Arg His Ser Cys Ala Glu Ala
 25                  30                  35                  40

TTG GTT GCC CTG CCG CCC GGA GCC TCA CCC CAG CGC TCC CGG AGC CCC      197
Leu Val Ala Leu Pro Pro Gly Ala Ser Pro Gln Arg Ser Arg Ser Pro
                 45                  50                  55

TCG CCG CAG CCC TCA TCT CAC GTG GCA CCC CAG GAC CAC GGC TCC CCG      245
Ser Pro Gln Pro Ser Ser His Val Ala Pro Gln Asp His Gly Ser Pro
             60                  65                  70

GCT GGG TAC CCC CCT GTG GCT GGC TCT GCC GTG ATC ATG GAT GCC CTG      293
Ala Gly Tyr Pro Pro Val Ala Gly Ser Ala Val Ile Met Asp Ala Leu
         75                  80                  85

AAC AGT CTC GCC ACG GAC TCG CCT TGT GGG ATC CCC CCC AAG ATG TGG      341
Asn Ser Leu Ala Thr Asp Ser Pro Cys Gly Ile Pro Pro Lys Met Trp
     90                  95                  100

AAGACC AGC CCT GAC CCC TCG CCG GTG TCT GCC GCC CCA TCC AAG GCC       389
Lys Thr Ser Pro Asp Pro Ser Pro Val Ser Ala Ala Pro Ser Lys Ala
105                 110                 115                 120

GGC CTG CCT CGC CAC ATC TAC CCG GCC GTG GAG TTC CTG GGG CCC TGC      437
Gly Leu Pro Arg His Ile Tyr Pro Ala Val Glu Phe Leu Gly Pro Cys
                125                 130                 135

GAG CAG GGC GAG AGG AGA AAC TCG GCT CCA GAA TCC ATC CTG CTG GTT      485
Glu Gln Gly Glu Arg Arg Asn Ser Ala Pro Glu Ser Ile Leu Leu Val
            140                 145                 150

CCG CCC ACT TGG CCC AAG CCG CTG GTG CCT GCC ATT CCC ATC TGC AGC      533
Pro Pro Thr Trp Pro Lys Pro Leu Val Pro Ala Ile Pro Ile Cys Ser
        155                 160                 165

ATC CCA GTG ACT GCA TCC CTC CCT CCA CTT GAG TGG CCG CTG TCC AGT      581
Ile Pro Val Thr Ala Ser Leu Pro Pro Leu Glu Trp Pro Leu Ser Ser
    170                 175                 180
```

```
CAG TCA GGC TCT TAC GAG CTG CGG ATC GAG GTG CAG CCC AAG CCA CAT      629
Gln Ser Gly Ser Tyr Glu Leu Arg Ile Glu Val Gln Pro Lys Pro His
185                 190                 195                 200

CAC CGG GCC CAC TAT GAG ACA GAA GGC AGC CGA GGG GCT GTC AAA GCT      677
His Arg Ala His Tyr Glu Thr Glu Gly Ser Arg Gly Ala Val Lys Ala
                205                 210                 215

CCA ACT GGA GGC CAC CCT GTG GTT CAG CTC CAT GGC TAC ATG GAA AAC      725
Pro Thr Gly Gly His Pro Val Val Gln Leu His Gly Tyr Met Glu Asn
            220                 225                 230

AAG CCT CTG GGA CTT CAG ATC TTC ATT GGG ACA GCT GAT GAG CGG ATC      773
Lys Pro Leu Gly Leu Gln Ile Phe Ile Gly Thr Ala Asp Glu Arg Ile
        235                 240                 245

CTT AAG CCG CAC GCC TTC TAC CAG GTG CAC CGA ATC ACG GGG AAA ACT      821
Leu Lys Pro His Ala Phe Tyr Gln Val His Arg Ile Thr Gly Lys Thr
    250                 255                 260

GTC ACC ACC ACC AGC TAT GAG AAG ATA GTG GGC AAC ACC AAA GTC CTG      869
Val Thr Thr Thr Ser Tyr Glu Lys Ile Val Gly Asn Thr Lys Val Leu
265                 270                 275                 280

GAG ATA CCC TTG GAG CCC AAA AAC AAC ATG AGG GCA ACC ATC GAC TGT      917
Glu Ile Pro Leu Glu Pro Lys Asn Asn Met Arg Ala Thr Ile Asp Cys
                285                 290                 295

GCG GGG ATC TTG AAG CTT AGA AAC GCC GAT ATT GAG CTG CGG AAA GGC      965
Ala Gly Ile Leu Lys Leu Arg Asn Ala Asp Ile Glu Leu Arg Lys Gly
            300                 305                 310

GAG ACG GAC ATT GGA AGA AAG AAC ACG CGG GTG AGA CTG GTT TTC CGA     1013
Glu Thr Asp Ile Gly Arg Lys Asn Thr Arg Val Arg Leu Val Phe Arg
        315                 320                 325

GTT CAC ATC CCA GAG TCC AGT GGC AGA ATC GTC TCT TTA CAG ACT GCA     1061
Val His Ile Pro Glu Ser Ser Gly Arg Ile Val Ser Leu Gln Thr Ala
    330                 335                 340

TCT AAC CCC ATC GAG TGC TCC CAG CGA TCT CGT CAC GAG CTG CCC ATG     1109
Ser Asn Pro Ile Glu Cys Ser Gln Arg Ser Arg His Glu Leu Pro Met
345                 350                 355                 360

GTT GAA AGA CAA GAC ACA GAC AGC TGC CTG GTC TAT GGC GGC CAG CAA     1157
Val Glu Arg Gln Asp Thr Asp Ser Cys Leu Val Tyr Gly Gly Gln Gln
                365                 370                 375

ATG ATC CTC ACG GGG CAG AAC TTT ACA TCC GAG TCC AAA GTT GTG TTT     1205
Met Ile Leu Thr Gly Gln Asn Phe Thr Ser Glu Ser Lys Val Val Phe
            380                 385                 390

ACT GAG AAG ACC ACA GAT GGA CAG CAA ATT TGG GAG ATG GAA GCC ACG     1253
Thr Glu Lys Thr Thr Asp Gly Gln Gln Ile Trp Glu Met Glu Ala Thr
        395                 400                 405
```

GTG GAT AAG GAC AAG AGC CAG CCC AAC ATG CTT TTT GTT GAG ATC CCT 1301
Val Asp Lys Asp Lys Ser Gln Pro Asn Met Leu Phe Val Glu Ile Pro
410 415 420

GAA TAT CGG AAC AAG CAT ATC CGC ACA CCT GTA AAA GTG AAC TTC TAC 1349
Glu Tyr Arg Asn Lys His Ile Arg Thr Pro Val Lys Val Asn Phe Tyr
425 430 435 440

GTC ATC AAT GGG AAG AGA AAA CGA AGT CAG CCT CAG CAC TTT ACC TAC 1397
Val Ile Asn Gly Lys Arg Lys Arg Ser Gln Pro Gln His Phe Thr Tyr
445 450 455

CAC CCA GTC CCA GCC ATC AAG ACG GAG CCC ACG GAT GAA TAT GAC CCC 1445
His Pro Val Pro Ala Ile Lys Thr Glu Pro Thr Asp Glu Tyr Asp Pro
460 465 470

ACT CTG ATC TGC AGC CCC ACC CAT GGA GGC CTG GGG AGC CAG CCT TAC 1493
Thr Leu Ile Cys Ser Pro Thr His Gly Gly Leu Gly Ser Gln Pro Tyr
475 480 485

TAC CCC CAG CAC CCG ATG GTG GCC GAG TCC CCC TCC TGC CTC GTG GCC 1541
Tyr Pro Gln His Pro Met Val Ala Glu Ser Pro Ser Cys Leu Val Ala
490 495 500

ACC ATG GCT CCC TGC CAG CAG TTC CGC ACG GGG CTC TCA TCC CCT GAC 1589
Thr Met Ala Pro Cys Gln Gln Phe Arg Thr Gly Leu Ser Ser Pro Asp
505 510 515 520

GCC CGC TAC CAG CAA CAG AAC CCA GCG GGC GTA CTC TAC CAG CGG AGC 1637
Ala Arg Tyr Gln Gln Gln Asn Pro Ala Gly Val Leu Tyr Gln Arg Ser
525 530 535

AAG AGC CTG AGC CCC AGC CTG CTG GGC TAT CAG CAG CCG GCC CTC ATG 1685
Lys Ser Leu Ser Pro Ser Leu Leu Gly Tyr Gln Gln Pro Ala Leu Met
540 545 550

GCC GCC CCG CTG TCC CTT GCG GAC GCT CAC CGC TCT GTG CTG GTG CAC 1733
Ala Ala Pro Leu Ser Leu Ala Asp Ala His Arg Ser Val Leu Val His
555 560 565

GCC GGC TCC CAG GGC CAG AGC TCA GCC CTG CTC CAC CCC TCT CCG ACC 1781
Ala Gly Ser Gln Gly Gln Ser Ser Ala Leu Leu His Pro Ser Pro Thr
570 575 580

AAC CAG AAG GCT TCG CCT GTG ATC CAC TAC TCA CCC ACC AAC CAG CAG 1829
Asn Gln Lys Ala Ser Pro Val Ile His Tyr Ser Pro Thr Asn Gln Gln
585 590 595 600

CTG CGC TGG GGA AGC CAC CAG GAG TTC CAG CAC ATC ATG TTC TGC GAG 1877
Leu Arg Trp Gly Ser His Gln Glu Phe Gln His Ile Met Phe Cys Glu
605 610 615

AAT TTC GCA CCA GGC ACC ACC AGA CCT GGC CCC CCC CCG GTC AGT CAA 1925
Asn Phe Ala Pro Gly Thr Thr Arg Pro Gly Pro Pro Pro Val Ser Gln
620 625 630

GGT CAG AGG CTG AGC CCG GGT TCC TAC CCC ACA GTC ATT CAG CAG CAG 1973
Gly Gln Arg Leu Ser Pro Gly Ser Tyr Pro Thr Val Ile Gln Gln Gln
635 640 645

AAT GCC ACG AGC CAA AGA GCC GCC AAA AAC GGA CCC CCG GTC AGT GAC 2021
Asn Ala Thr Ser Gln Arg Ala Ala Lys Asn Gly Pro Pro Val Ser Asp
650 655 660

CAA AAG GAA GTA TTA CCT GCG GGG GTG ACC ATT AAA CAG GAG CAG AAC 2069
Gln Lys Glu Val Leu Pro Ala Gly Val Thr Ile Lys Gln Glu Gln Asn
665 670 675 680

TTG GAG GCC AAC CCT GGA GTG GAA GTC CTC TGT GGG GTA GGG ATG GCT 2117
Leu Glu Ala Asn Pro Gly Val Glu Val Leu Cys Gly Val Gly Met Ala
685 690 695

GAT GGG GCA TAGCACTTGT TAGGGGCTGA AGCAGAAGAG TCAGAGTTCT 2166
Asp Gly Ala
700

GAAGAGTAAG AGAAGATAGT GAAGCCAGCC CACTTGTGAC AGCAGAGGAT AAAGCAGAGG 2226

AGTTGATTAA ATTGGTGTCA TTGGATGTCA GAAAACCTTT TAACGCAGAC AAAAGAGGAC 2286

TGTTCACACC AAGTGGACCG GCAACGCTGG GAGTAGAGCC ACCAGCAATT ACAGGAGTCG 2346

GGTTGGATAG GCCTTGGGAT GTTGGTGGTA AGGGCAAAGG GAGCCCTGCA AAAACTGATG 2406

ACAATGAAGC AGAATTTGGG TTGCTGGTAG AAGCAGCAGA AGAGCTGGTG GAAAAGGGGA 2466

GGCTAGTGAA AGGTGCAGAA GTAGAAGCAA ATGCTTCACT GGAACCAAGA GTGGACCGTG 2526

GTGTAGGTCC TGGGGTAGGA GTGGCTGCGG TAGGAGTGGC AGAAGGACCT GGCAGTGACA 2586

CTAGGCCAGA AAAAATGGAA GGAACAGGAG TGGTGGTTGT ACTGTGGATG GATGTAACAG 2646

GTGCAGTGGG CAATGAAGGG GTTCTGATAA CTGTTGGGTT TGGTATTGAT GTCTGAGGTG 2706

TGTGAACGGC TGAGGAGACC TCCCCTGGGA ACACAGGAAG GACAGTATTC AGCAGGTTCA 2766

TTCCAGAAAC GGTGGCACCT GCTGATGCTG ATGGATGATT AATTCCCTTG ACTGGGGATA 2826

CAGTAGGAAC AGG 2839

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 699 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

```
Met Ser Pro Arg Thr Ser Leu Ala Asp Asp Ser Cys Leu Gly Arg His
 1               5                  10                  15

Ser Pro Val Pro Arg Pro Ala Ser Arg Ser Ser Ser Pro Gly Ala Lys
            20                  25                  30

Arg Arg His Ser Cys Ala Glu Ala Leu Val Ala Leu Pro Pro Gly Ala
        35                  40                  45

Ser Pro Gln Arg Ser Arg Ser Pro Ser Pro Gln Pro Ser Ser His Val
    50                  55                  60

Ala Pro Gln Asp His Gly Ser Pro Ala Gly Tyr Pro Pro Val Ala Gly
65                  70                  75                  80

Ser Ala Val Ile Met Asp Ala Leu Asn Ser Leu Ala Thr Asp Ser Pro
                85                  90                  95

Cys Gly Ile Pro Pro Lys Met Trp Lys Thr Ser Pro Asp Pro Ser Pro
            100                 105                 110

Val Ser Ala Ala Pro Ser Lys Ala Gly Leu Pro Arg His Ile Tyr Pro
        115                 120                 125

Ala Val Glu Phe Leu Gly Pro Cys Glu Gln Gly Glu Arg Arg Asn Ser
    130                 135                 140

Ala Pro Glu Ser Ile Leu Leu Val Pro Pro Thr Trp Pro Lys Pro Leu
145                 150                 155                 160

Val Pro Ala Ile Pro Ile Cys Ser Ile Pro Val Thr Ala Ser Leu Pro
                165                 170                 175

Pro Leu Glu Trp Pro Leu Ser Ser Gln Ser Gly Ser Tyr Glu Leu Arg
            180                 185                 190

Ile Glu Val Gln Pro Lys Pro His His Arg Ala His Tyr Glu Thr Glu
        195                 200                 205

Gly Ser Arg Gly Ala Val Lys Ala Pro Thr Gly Gly His Pro Val Val
    210                 215                 220

Gln Leu His Gly Tyr Met Glu Asn Lys Pro Leu Gly Leu Gln Ile Phe
225                 230                 235                 240

Ile Gly Thr Ala Asp Glu Arg Ile Leu Lys Pro His Ala Phe Tyr Gln
                245                 250                 255

Val His Arg Ile Thr Gly Lys Thr Val Thr Thr Thr Ser Tyr Glu Lys
            260                 265                 270

Ile Val Gly Asn Thr Lys Val Leu Glu Ile Pro Leu Glu Pro Lys Asn
        275                 280                 285
```

Asn Met Arg Ala Thr Ile Asp Cys Ala Gly Ile Leu Lys Leu Arg Asn
290 295 300

Ala Asp Ile Glu Leu Arg Lys Gly Glu Thr Asp Ile Gly Arg Lys Asn
305 310 315 320

Thr Arg Val Arg Leu Val Phe Arg Val His Ile Pro Glu Ser Ser Gly
325 330 335

Arg Ile Val Ser Leu Gln Thr Ala Ser Asn Pro Ile Glu Cys Ser Gln
340 345 350

Arg Ser Arg His Glu Leu Pro Met Val Glu Arg Gln Asp Thr Asp Ser
355 360 365

Cys Leu Val Tyr Gly Gly Gln Gln Met Ile Leu Thr Gly Gln Asn Phe
370 375 380

Thr Ser Glu Ser Lys Val Val Phe Thr Glu Lys Thr Thr Asp Gly Gln
385 390 395 400

Gln Ile Trp Glu Met Glu Ala Thr Val Asp Lys Asp Lys Ser Gln Pro
405 410 415

Asn Met Leu Phe Val Glu Ile Pro Glu Tyr Arg Asn Lys His Ile Arg
420 425 430

Thr Pro Val Lys Val Asn Phe Tyr Val Ile Asn Gly Lys Arg Lys Arg
435 440 445

Ser Gln Pro Gln His Phe Thr Tyr His Pro Val Pro Ala Ile Lys Thr
450 455 460

Glu Pro Thr Asp Glu Tyr Asp Pro Thr Leu Ile Cys Ser Pro Thr His
465 470 475 480

Gly Gly Leu Gly Ser Gln Pro Tyr Tyr Pro Gln His Pro Met Val Ala
485 490 495

Glu Ser Pro Ser Cys Leu Val Ala Thr Met Ala Pro Cys Gln Gln Phe
500 505 510

Arg Thr Gly Leu Ser Ser Pro Asp Ala Arg Tyr Gln Gln Gln Asn Pro
515 520 525

Ala Gly Val Leu Tyr Gln Arg Ser Lys Ser Leu Ser Pro Ser Leu Leu
530 535 540

Gly Tyr Gln Gln Pro Ala Leu Met Ala Ala Pro Leu Ser Leu Ala Asp
545 550 555 560

Ala His Arg Ser Val Leu Val His Ala Gly Ser Gln Gly Gln Ser Ser
565 570 575

Ala Leu Leu His Pro Ser Pro Thr Asn Gln Lys Ala Ser Pro Val Ile
580 585 590

His Tyr Ser Pro Thr Asn Gln Gln Leu Arg Trp Gly Ser His Gln Glu
595 600 605

Phe Gln His Ile Met Phe Cys Glu Asn Phe Ala Pro Gly Thr Thr Arg
610 615 620

Pro Gly Pro Pro Pro Val Ser Gln Gly Gln Arg Leu Ser Pro Gly Ser
625 630 635 640

Tyr Pro Thr Val Ile Gln Gln Gln Asn Ala Thr Ser Gln Arg Ala Ala
645 650 655

Lys Asn Gly Pro Pro Val Ser Asp Gln Lys Glu Val Leu Pro Ala Gly
660 665 670

Val Thr Ile Lys Gln Glu Gln Asn Leu Glu Ala Asn Pro Gly Val Glu
675 680 685

Val Leu Cys Gly Val Gly Met Ala Asp Gly Ala
690 695

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4010 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 304..3531

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: replace(1756..1758, "gta")

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: replace(1756..1758, "gaa")

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: replace(1756..1758, "gga")

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: replace(3090..3092, "agt")

(ix) FEATURE:
(A) NAME/KEY: unsure

(B) LOCATION: replace(3090..3092, "aga")

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: replace(3090..3092, "agg")

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

```
GAGCGCACCC GCGGCGGCGT TGGCGGCGAC TGTGGGGGGG CGGCGGGGAA CATTGGCTAA     60

GCCGACAGTG GAGGCTTAGG CACCGGTGGC GGGCGGCTGC GGTTCCTGGT GCTGCTCGGC    120

GCGCGGCCAG CTTTCGGAAC GGAACGCTCG GCGTCGCGGG CCCCGCCCGG AAAGTTTGCC    180

GTGGAGTCGC GACCTCTTGG CCCGCGCGGC CCGGAATTAA GCGGGGTTGA GGAGCTGTTG    240

TCGCCGCTTG CCGTTGCCGC CGCCGCCGCC TGAAGAGGAG CTGCAGCACC CTGGGCCACG    300

CCG ATG ACT ACT GCA AAC TGT GGC GCC CAC GAC GAG CTC GAC TTC AAA       348
    Met Thr Thr Ala Asn Cys Gly Ala His Asp Glu Leu Asp Phe Lys
      1               5                  10                  15

CTC GTC TTT GGC GAG GAC GGG GCG CCG GCG CCG CCG CCC CCG GGC TCG      396
Leu Val Phe Gly Glu Asp Gly Ala Pro Ala Pro Pro Pro Pro Gly Ser
                  20                  25                  30

CGG CCT GCA GAT CTT GAG CCA GAT GAT TGT GCA TCC ATT TAC ATC TTT      444
Arg Pro Ala Asp Leu Glu Pro Asp Asp Cys Ala Ser Ile Tyr Ile Phe
              35                  40                  45

AAT GTA GAT CCA CCT CCA TCT ACT TTA ACC ACA CCA CTT TGC TTA CCA      492
Asn Val Asp Pro Pro Pro Ser Thr Leu Thr Thr Pro Leu Cys Leu Pro
          50                  55                  60

CAT CAT GGA TTA CCG TCT CAC TCT TCT GTT TTG TCA CCA TCG TTT CAG      540
His His Gly Leu Pro Ser His Ser Ser Val Leu Ser Pro Ser Phe Gln
      65                  70                  75

CTC CAA AGT CAC AAA AAC TAT GAA GGA ACT TGT GAG ATT CCT GAA TCT      588
Leu Gln Ser His Lys Asn Tyr Glu Gly Thr Cys Glu Ile Pro Glu Ser
  80                  85                  90                  95

AAA TAT AGC CCA TTA GGT GGT CCC AAA CCC TTT GAG TGC CCA AGT ATT      636
Lys Tyr Ser Pro Leu Gly Gly Pro Lys Pro Phe Glu Cys Pro Ser Ile
              100                 105                 110

CAA ATT ACA TCT ATC TCT CCT AAC TGT CAT CAA GAA TTA GAT GCA CAT      684
Gln Ile Thr Ser Ile Ser Pro Asn Cys His Gln Glu Leu Asp Ala His
          115                 120                 125

GAA GAT GAC CTA CAG ATA AAT GAC CCA GAA CGG GAA TTT TTG GAA AGG      732
Glu Asp Asp Leu Gln Ile Asn Asp Pro Glu Arg Glu Phe Leu Glu Arg
      130                 135                 140
```

```
CCT TCT AGA GAT CAT CTC TAT CTT CCT CTT GAG CCA TCC TAC CGG GAG      780
Pro Ser Arg Asp His Leu Tyr Leu Pro Leu Glu Pro Ser Tyr Arg Glu
    145                 150                 155

TCT TCT CTT AGT CCT AGT CCT GCC AGC AGC ATC TCT TCT AGG AGT TGG      828
Ser Ser Leu Ser Pro Ser Pro Ala Ser Ser Ile Ser Ser Arg Ser Trp
160                 165                 170                 175

TTC TCT GAT GCA TCT TCT TGT GAA TCG CTT TCA CAT ATT TAT GAT GAT      876
Phe Ser Asp Ala Ser Ser Cys Glu Ser Leu Ser His Ile Tyr Asp Asp
                180                 185                 190

GTG GAC TCA GAG TTG AAT GAA GCT GCA GCC CGA TTT ACC CTT GGA TCC      924
Val Asp Ser Glu Leu Asn Glu Ala Ala Ala Arg Phe Thr Leu Gly Ser
            195                 200                 205

CCT CTG ACT TCT CCT GGT GGC TCT CCA GGG GGC TGC CCT GGA GAA GAA      972
Pro Leu Thr Ser Pro Gly Gly Ser Pro Gly Gly Cys Pro Gly Glu Glu
        210                 215                 220

ACT TGG CAT CAA CAG TAT GGA CTT GGA CAC TCA TTA TCA CCC AGG CAA     1020
Thr Trp His Gln Gln Tyr Gly Leu Gly His Ser Leu Ser Pro Arg Gln
    225                 230                 235

TCT CCT TGC CAC TCT CCT AGA TCC AGT GTC ACT GAT GAG AAT TGG CTG     1068
Ser Pro Cys His Ser Pro Arg Ser Ser Val Thr Asp Glu Asn Trp Leu
240                 245                 250                 255

AGC CCC AGG CCA GCC TCA GGA CCC TCA TCA AGG CCC ACA TCC CCC TGT     1116
Ser Pro Arg Pro Ala Ser Gly Pro Ser Ser Arg Pro Thr Ser Pro Cys
                260                 265                 270

GGG AAA CGG AGG CAC TCC AGT GCT GAA GTT TGT TAT GCT GGG TCC CTT     1164
Gly Lys Arg Arg His Ser Ser Ala Glu Val Cys Tyr Ala Gly Ser Leu
            275                 280                 285

TCA CCC CAT CAC TCA CCT GTT CCT TCA CCT GGT CAC TCC CCC AGG GGA     1212
Ser Pro His His Ser Pro Val Pro Ser Pro Gly His Ser Pro Arg Gly
        290                 295                 300

AGT GTG ACA GAA GAT ACG TGG CTC AAT GCT TCT GTC CAT GGT GGG TCA     1260
Ser Val Thr Glu Asp Thr Trp Leu Asn Ala Ser Val His Gly Gly Ser
    305                 310                 315

GGC CTT GGC CCT GCA GTT TTT CCA TTT CAG TAC TGT GTA GAG ACT GAC     1308
Gly Leu Gly Pro Ala Val Phe Pro Phe Gln Tyr Cys Val Glu Thr Asp
320                 325                 330                 335

ATC CCT CTC AAA ACA AGG AAA ACT TCT GAA GAT CAA GCT GCC ATA CTA     1356
Ile Pro Leu Lys Thr Arg Lys Thr Ser Glu Asp Gln Ala Ala Ile Leu
                340                 345                 350

CCA GGA AAA TTA GAG CTG TGT TCA GAT GAC CAA GGG AGT TTA TCA CCA     1404
Pro Gly Lys Leu Glu Leu Cys Ser Asp Asp Gln Gly Ser Leu Ser Pro
            355                 360                 365
```

```
GCC CGG GAG ACT TCA ATA GAT GAT GGC CTT GGA TCT CAG TAT CCT TTA      1452
Ala Arg Glu Thr Ser Ile Asp Asp Gly Leu Gly Ser Gln Tyr Pro Leu
        370                 375                 380

AAG AAA GAT TCA TGT GGT GAT CAG TTT CTT TCA GTT CCT TCA CCC TTT      1500
Lys Lys Asp Ser Cys Gly Asp Gln Phe Leu Ser Val Pro Ser Pro Phe
    385                 390                 395

ACC TGG AGC AAA CCA AAG CCT GGC CAC ACC CCT ATA TTT CGC ACA TCT      1548
Thr Trp Ser Lys Pro Lys Pro Gly His Thr Pro Ile Phe Arg Thr Ser
400                 405                 410                 415

TCA TTA CCT CCA CTA GAC TGG CCT TTA CCA GCT CAT TTT GGA CAA TGT      1596
Ser Leu Pro Pro Leu Asp Trp Pro Leu Pro Ala His Phe Gly Gln Cys
                420                 425                 430

GAA CTG AAA ATA GAA GTG CAA CCT AAA ACT CAT CAT CGA GCC CAT TAT      1644
Glu Leu Lys Ile Glu Val Gln Pro Lys Thr His His Arg Ala His Tyr
            435                 440                 445

GAA ACT GAA GGT AGC CGA GGG GCA GTA AAA GCA TCT ACT GGG GGA CAT      1692
Glu Thr Glu Gly Ser Arg Gly Ala Val Lys Ala Ser Thr Gly Gly His
        450                 455                 460

CCT GTT GTG AAG CTC CTG GGC TAT AAC GAA AAG CCA ATA AAT CTA CAA      1740
Pro Val Val Lys Leu Leu Gly Tyr Asn Glu Lys Pro Ile Asn Leu Gln
    465                 470                 475

ATG TTT ATT GGG ACA GCA GAT GAT CGA TAT TTA CGA CCT CAT GCA TTT      1788
Met Phe Ile Gly Thr Ala Asp Asp Arg Tyr Leu Arg Pro His Ala Phe
480                 485                 490                 495

TAC CAG GTG CAT CGA ATC ACT GGG AAG ACA GTC GCT ACT GCA AGC CAA      1836
Tyr Gln Val His Arg Ile Thr Gly Lys Thr Val Ala Thr Ala Ser Gln
                500                 505                 510

GAG ATA ATA ATT GCC AGT ACA AAA GTT CTG GAA ATT CCA CTT CTT CCT      1884
Glu Ile Ile Ile Ala Ser Thr Lys Val Leu Glu Ile Pro Leu Leu Pro
            515                 520                 525

GAA AAT AAT ATG TCA GCC AGT ATT GAT TGT GCA GGT ATT TTG AAA CTC      1932
Glu Asn Asn Met Ser Ala Ser Ile Asp Cys Ala Gly Ile Leu Lys Leu
        530                 535                 540

CGC AAT TCA GAT ATA GAA CTT CGA AAA GGA GAA ACT GAT ATT GGC AGA      1980
Arg Asn Ser Asp Ile Glu Leu Arg Lys Gly Glu Thr Asp Ile Gly Arg
    545                 550                 555

AAG AAT ACT AGA GTA CGA CTT GTG TTT CGT GTA CAC ATC CCA CAG CCC      2028
Lys Asn Thr Arg Val Arg Leu Val Phe Arg Val His Ile Pro Gln Pro
560                 565                 570                 575

AGT GGA AAA GTC CTT TCT CTG CAG ATA GCC TCT ATA CCC GTT GAG TGC      2076
Ser Gly Lys Val Leu Ser Leu Gln Ile Ala Ser Ile Pro Val Glu Cys
                580                 585                 590
```

TCC CAG CGG TCT GCT CAA GAA CTT CCT CAT ATT GAG AAG TAC AGT ATC 2124
Ser Gln Arg Ser Ala Gln Glu Leu Pro His Ile Glu Lys Tyr Ser Ile
595 600 605

AAC AGT TGT TCT GTA AAT GGA GGT CAT GAA ATG GTT GTG ACT GGA TCT 2172
Asn Ser Cys Ser Val Asn Gly Gly His Glu Met Val Val Thr Gly Ser
610 615 620

AAT TTT CTT CCA GAA TCC AAA ATC ATT TTT CTT GAA AAA GGA CAA GAT 2220
Asn Phe Leu Pro Glu Ser Lys Ile Ile Phe Leu Glu Lys Gly Gln Asp
625 630 635

GGA CGA CCT CAG TGG GAG GTA GAA GGG AAG ATA ATC AGG GAA AAA TGT 2268
Gly Arg Pro Gln Trp Glu Val Glu Gly Lys Ile Ile Arg Glu Lys Cys
640 645 650 655

CAA GGG GCT CAC ATT GTC CTT GAA GTT CCT CCA TAT CAT AAC CCA GCA 2316
Gln Gly Ala His Ile Val Leu Glu Val Pro Pro Tyr His Asn Pro Ala
660 665 670

GTT ACA GCT GCA GTG CAG GTG CAC TTT TAT CTT TGC AAT GGC AAG AGG 2364
Val Thr Ala Ala Val Gln Val His Phe Tyr Leu Cys Asn Gly Lys Arg
675 680 685

AAA AAA AGC CAG TCT CAA CGT TTT ACT TAT ACA CCA GTT TTG CTG AAG 2412
Lys Lys Ser Gln Ser Gln Arg Phe Thr Tyr Thr Pro Val Leu Leu Lys
690 695 700

CAA GAA CAC AGA GAA GAG ATT GAT TTG TCT TCA GTT CCA TCT TTG CCT 2460
Gln Glu His Arg Glu Glu Ile Asp Leu Ser Ser Val Pro Ser Leu Pro
705 710 715

GTG CCT CAT CCT GCT CAG ACC CAG AGG CCT TCC TCT GAT TCA GGG TGT 2508
Val Pro His Pro Ala Gln Thr Gln Arg Pro Ser Ser Asp Ser Gly Cys
720 725 730 735

TCA CAT GAC AGT GTA CTG TCA GGA CAG AGA AGT TTG ATT TGC TCC ATC 2556
Ser His Asp Ser Val Leu Ser Gly Gln Arg Ser Leu Ile Cys Ser Ile
740 745 750

CCA CAA ACA TAT GCA TCC ATG GTG ACC TCA TCC CAT CTG CCA CAG TTG 2604
Pro Gln Thr Tyr Ala Ser Met Val Thr Ser Ser His Leu Pro Gln Leu
755 760 765

CAG TGT AGA GAT GAG AGT GTT AGT AAA GAA CAG CAT ATG ATT CCT TCT 2652
Gln Cys Arg Asp Glu Ser Val Ser Lys Glu Gln His Met Ile Pro Ser
770 775 780

CCA ATT GTA CAC CAG CCT TTT CAA GTC ACA CCA ACA CCT CCT GTG GGG 2700
Pro Ile Val His Gln Pro Phe Gln Val Thr Pro Thr Pro Pro Val Gly
785 790 795

TCT TCC TAT CAG CCT ATG CAA ACT AAT GTT GTG TAC AAT GGA CCA ACT 2748
Ser Ser Tyr Gln Pro Met Gln Thr Asn Val Val Tyr Asn Gly Pro Thr
800 805 810 815

```
TGT CTT CCT ATT AAT GCT GCC TCT AGT CAA GAA TTT GAT TCA GTT TGG      2796
Cys Leu Pro Ile Asn Ala Ala Ser Ser Gln Glu Phe Asp Ser Val Trp
                820                 825                 830

TTT CAG CAG GAT GCA ACT CTT TCT GGT TTA GTG AAT CTT GGC TGT CAA      2844
Phe Gln Gln Asp Ala Thr Leu Ser Gly Leu Val Asn Leu Gly Cys Gln
            835                 840                 845

CCA CTG TCA TCC ATA CCA TTT CAT TCT TCA AAT TCA GGC TCA ACA GGA      2892
Pro Leu Ser Ser Ile Pro Phe His Ser Ser Asn Ser Gly Ser Thr Gly
        850                 855                 860

CAT CTC TTA GCC CAT ACA CCT CAT TCT GTG CAT ACC CTG CCT CAT CTG      2940
His Leu Leu Ala His Thr Pro His Ser Val His Thr Leu Pro His Leu
    865                 870                 875

CAA TCA ATG GGA TAT CAT TGT TCA AAT ACA GGA CAA AGA TCT CTT TCT      2988
Gln Ser Met Gly Tyr His Cys Ser Asn Thr Gly Gln Arg Ser Leu Ser
880                 885                 890                 895

TCT CCA GTG GGT GAC CAG ATT ACA GGT CAG CCT TCG TCT CAG TTA CAA      3036
Ser Pro Val Gly Asp Gln Ile Thr Gly Gln Pro Ser Ser Gln Leu Gln
                900                 905                 910

CCT ATT ACA TAT GGT CCT TCA CAT TCA GGG TCT GTT ACA ACA GCT TCC      3084
Pro Ile Thr Tyr Gly Pro Ser His Ser Gly Ser Val Thr Thr Ala Ser
            915                 920                 925

CCA GCA GCT TCT CAT CCC TTG GGT AGT TCA CCG CTT TCT GGG CCA CCA      3132
Pro Ala Ala Ser His Pro Leu Gly Ser Ser Pro Leu Ser Gly Pro Pro
        930                 935                 940

TCT CCT CAG TTT CAG CCT ATG CCT TAC CAA TCT CCT AGC TCA GGA ACT      3180
Ser Pro Gln Phe Gln Pro Met Pro Tyr Gln Ser Pro Ser Ser Gly Thr
    945                 950                 955

GGC TCA TCA CCG TCT CCA GCC ACC AGA ATG CAT TCT GGA CAG CAC TCA      3228
Gly Ser Ser Pro Ser Pro Ala Thr Arg Met His Ser Gly Gln His Ser
960                 965                 970                 975

ACT CAA GCA CAA AGT ACG GGC CAG GGG GGT CTT TCT GCA CCT TCA TCC      3276
Thr Gln Ala Gln Ser Thr Gly Gln Gly Gly Leu Ser Ala Pro Ser Ser
                980                 985                 990

TTA ATA TGT CAC AGT TTG TGT GAT CCA GCG TCA TTT CCA CCT GAT GGG      3324
Leu Ile Cys His Ser Leu Cys Asp Pro Ala Ser Phe Pro Pro Asp Gly
            995                 1000                1005

GCA ACT GTG AGC ATT AAA CCT GAA CCA GAA GAT CGA GAG CCT AAC TTT      3372
Ala Thr Val Ser Ile Lys Pro Glu Pro Glu Asp Arg Glu Pro Asn Phe
        1010                1015                1020

GCA ACC ATT GGT CTG CAG GAC ATC ACT TTA GAT GAT GTG AAC GAG ATA      3420
Ala Thr Ile Gly Leu Gln Asp Ile Thr Leu Asp Asp Val Asn Glu Ile
    1025                1030                1035
```

```
ATT GGG AGA GAC ATG TCC CAG ATT TCT GTT TCC CAA GGA GCA GGG GTG      3468
Ile Gly Arg Asp Met Ser Gln Ile Ser Val Ser Gln Gly Ala Gly Val
1040                1045                1050                1055

AGC AGG CAG GCT CCC CTC CCG AGT CCT GAG TCC CTG GAT TTA GGA AGA      3516
Ser Arg Gln Ala Pro Leu Pro Ser Pro Glu Ser Leu Asp Leu Gly Arg
                    1060                1065                1070

TCT GAT GGG CTC TAACAGTGCT TACTGCAGCC TTGTGTCCAC CACCAACTTC          3568
Ser Asp Gly Leu
            1075

TCAGCATGTT TCTCTCCTTG GACCTTGGGT TTCCAACTCT TCAACCTTCA GGTCTGGGGC   3628

CAGGAGTGGG ACCCACCATT TGTGGGGAAA GTAGCATTCC TCCACCTCAG GCCTTGGGTA   3688

GATTTGGCAA AAGAACAGGA GCAGCATAGG CTGTTTGAGC TTTGGGGAAA TGAACTTTGC   3748

TTTTTATATT TAACTAGGAT ACTTTTAAAT GATGGGTGCT TTGAGTGTGA ATCCAGCAGG   3808

CTCTCTTGTT TCCGAGGTGC TGCTTTTGCA GGTGACCTGG TTACTTAACT AGGAGTGGTG   3868

ATTTGTACTG CTTTATGGTC ATTTGAAGGG CCCCTTAGTT TTTATGATAA TTTTTAAAAT   3928

AGGAACTTTT GATAAGACCT TCTAGAACCC CAAAAAAAAA AAAAAAAGAA AAAAAAAGAA   3988

AAACAATAAA AAAAAAAAAA GG                                            4010
```

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1075 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

```
Met Thr Thr Ala Asn Cys Gly Ala His Asp Glu Leu Asp Phe Lys Leu
 1               5                  10                  15

Val Phe Gly Glu Asp Gly Ala Pro Ala Pro Pro Pro Pro Gly Ser Arg
            20                  25                  30

Pro Ala Asp Leu Glu Pro Asp Asp Cys Ala Ser Ile Tyr Ile Phe Asn
        35                  40                  45

Val Asp Pro Pro Pro Ser Thr Leu Thr Thr Pro Leu Cys Leu Pro His
    50                  55                  60

His Gly Leu Pro Ser His Ser Ser Val Leu Ser Pro Ser Phe Gln Leu
65                  70                  75                  80
```

Gln Ser His Lys Asn Tyr Glu Gly Thr Cys Glu Ile Pro Glu Ser Lys
85 90 95

Tyr Ser Pro Leu Gly Gly Pro Lys Pro Phe Glu Cys Pro Ser Ile Gln
100 105 110

Ile Thr Ser Ile Ser Pro Asn Cys His Gln Glu Leu Asp Ala His Glu
115 120 125

Asp Asp Leu Gln Ile Asn Asp Pro Glu Arg Glu Phe Leu Glu Arg Pro
130 135 140

Ser Arg Asp His Leu Tyr Leu Pro Leu Glu Pro Ser Tyr Arg Glu Ser
145 150 155 160

Ser Leu Ser Pro Ser Pro Ala Ser Ser Ile Ser Ser Arg Ser Trp Phe
165 170 175

Ser Asp Ala Ser Ser Cys Glu Ser Leu Ser His Ile Tyr Asp Asp Val
180 185 190

Asp Ser Glu Leu Asn Glu Ala Ala Ala Arg Phe Thr Leu Gly Ser Pro
195 200 205

Leu Thr Ser Pro Gly Gly Ser Pro Gly Gly Cys Pro Gly Glu Glu Thr
210 215 220

Trp His Gln Gln Tyr Gly Leu Gly His Ser Leu Ser Pro Arg Gln Ser
225 230 235 240

Pro Cys His Ser Pro Arg Ser Ser Val Thr Asp Glu Asn Trp Leu Ser
245 250 255

Pro Arg Pro Ala Ser Gly Pro Ser Ser Arg Pro Thr Ser Pro Cys Gly
260 265 270

Lys Arg Arg His Ser Ser Ala Glu Val Cys Tyr Ala Gly Ser Leu Ser
275 280 285

Pro His His Ser Pro Val Pro Ser Pro Gly His Ser Pro Arg Gly Ser
290 295 300

Val Thr Glu Asp Thr Trp Leu Asn Ala Ser Val His Gly Gly Ser Gly
305 310 315 320

Leu Gly Pro Ala Val Phe Pro Phe Gln Tyr Cys Val Glu Thr Asp Ile
325 330 335

Pro Leu Lys Thr Arg Lys Thr Ser Glu Asp Gln Ala Ala Ile Leu Pro
340 345 350

Gly Lys Leu Glu Leu Cys Ser Asp Asp Gln Gly Ser Leu Ser Pro Ala
355 360 365

Arg Glu Thr Ser Ile Asp Asp Gly Leu Gly Ser Gln Tyr Pro Leu Lys
370 375 380

Lys Asp Ser Cys Gly Asp Gln Phe Leu Ser Val Pro Ser Pro Phe Thr
385 390 395 400

Trp Ser Lys Pro Lys Pro Gly His Thr Pro Ile Phe Arg Thr Ser Ser
405 410 415

Leu Pro Pro Leu Asp Trp Pro Leu Pro Ala His Phe Gly Gln Cys Glu
420 425 430

Leu Lys Ile Glu Val Gln Pro Lys Thr His His Arg Ala His Tyr Glu
435 440 445

Thr Glu Gly Ser Arg Gly Ala Val Lys Ala Ser Thr Gly Gly His Pro
450 455 460

Val Val Lys Leu Leu Gly Tyr Asn Glu Lys Pro Ile Asn Leu Gln Met
465 470 475 480

Phe Ile Gly Thr Ala Asp Asp Arg Tyr Leu Arg Pro His Ala Phe Tyr
485 490 495

Gln Val His Arg Ile Thr Gly Lys Thr Val Ala Thr Ala Ser Gln Glu
500 505 510

Ile Ile Ile Ala Ser Thr Lys Val Leu Glu Ile Pro Leu Leu Pro Glu
515 520 525

Asn Asn Met Ser Ala Ser Ile Asp Cys Ala Gly Ile Leu Lys Leu Arg
530 535 540

Asn Ser Asp Ile Glu Leu Arg Lys Gly Glu Thr Asp Ile Gly Arg Lys
545 550 555 560

Asn Thr Arg Val Arg Leu Val Phe Arg Val His Ile Pro Gln Pro Ser
565 570 575

Gly Lys Val Leu Ser Leu Gln Ile Ala Ser Ile Pro Val Glu Cys Ser
580 585 590

Gln Arg Ser Ala Gln Glu Leu Pro His Ile Glu Lys Tyr Ser Ile Asn
595 600 605

Ser Cys Ser Val Asn Gly Gly His Glu Met Val Val Thr Gly Ser Asn
610 615 620

Phe Leu Pro Glu Ser Lys Ile Ile Phe Leu Glu Lys Gly Gln Asp Gly
625 630 635 640

Arg Pro Gln Trp Glu Val Glu Gly Lys Ile Ile Arg Glu Lys Cys Gln
645 650 655

Gly Ala His Ile Val Leu Glu Val Pro Pro Tyr His Asn Pro Ala Val
660 665 670

Thr Ala Ala Val Gln Val His Phe Tyr Leu Cys Asn Gly Lys Arg Lys
675 680 685

Lys Ser Gln Ser Gln Arg Phe Thr Tyr Thr Pro Val Leu Leu Lys Gln
690 695 700

Glu His Arg Glu Glu Ile Asp Leu Ser Ser Val Pro Ser Leu Pro Val
705 710 715 720

Pro His Pro Ala Gln Thr Gln Arg Pro Ser Ser Asp Ser Gly Cys Ser
725 730 735

His Asp Ser Val Leu Ser Gly Gln Arg Ser Leu Ile Cys Ser Ile Pro
740 745 750

Gln Thr Tyr Ala Ser Met Val Thr Ser Ser His Leu Pro Gln Leu Gln
755 760 765

Cys Arg Asp Glu Ser Val Ser Lys Glu Gln His Met Ile Pro Ser Pro
770 775 780

Ile Val His Gln Pro Phe Gln Val Thr Pro Thr Pro Pro Val Gly Ser
785 790 795 800

Ser Tyr Gln Pro Met Gln Thr Asn Val Val Tyr Asn Gly Pro Thr Cys
805 810 815

Leu Pro Ile Asn Ala Ala Ser Ser Gln Glu Phe Asp Ser Val Trp Phe
820 825 830

Gln Gln Asp Ala Thr Leu Ser Gly Leu Val Asn Leu Gly Cys Gln Pro
835 840 845

Leu Ser Ser Ile Pro Phe His Ser Ser Asn Ser Gly Ser Thr Gly His
850 855 860

Leu Leu Ala His Thr Pro His Ser Val His Thr Leu Pro His Leu Gln
865 870 875 880

Ser Met Gly Tyr His Cys Ser Asn Thr Gly Gln Arg Ser Leu Ser Ser
885 890 895

Pro Val Gly Asp Gln Ile Thr Gly Gln Pro Ser Ser Gln Leu Gln Pro
900 905 910

Ile Thr Tyr Gly Pro Ser His Ser Gly Ser Val Thr Thr Ala Ser Pro
915 920 925

Ala Ala Ser His Pro Leu Gly Ser Ser Pro Leu Ser Gly Pro Pro Ser
930 935 940

Pro Gln Phe Gln Pro Met Pro Tyr Gln Ser Pro Ser Ser Gly Thr Gly
945 950 955 960

Ser Ser Pro Ser Pro Ala Thr Arg Met His Ser Gly Gln His Ser Thr
965 970 975

Gln Ala Gln Ser Thr Gly Gln Gly Gly Leu Ser Ala Pro Ser Ser Leu
980 985 990

Ile Cys His Ser Leu Cys Asp Pro Ala Ser Phe Pro Pro Asp Gly Ala
995 1000 1005

Thr Val Ser Ile Lys Pro Glu Pro Glu Asp Arg Glu Pro Asn Phe Ala
1010 1015 1020

Thr Ile Gly Leu Gln Asp Ile Thr Leu Asp Asp Val Asn Glu Ile Ile
1025 1030 1035 1040

Gly Arg Asp Met Ser Gln Ile Ser Val Ser Gln Gly Ala Gly Val Ser
1045 1050 1055

Arg Gln Ala Pro Leu Pro Ser Pro Glu Ser Leu Asp Leu Gly Arg Ser
1060 1065 1070

Asp Gly Leu
1075

CLAIMS:

1. A composition selected from the group consisting of:
   a) a substantially pure NF-AT120 protein or peptide thereof, or a fusion protein comprising NF-AT120 protein sequence;
   b) an antibody specific for binding to an NF-AT120 protein; and
   c) a nucleic acid encoding an NF-AT120 protein or fragment thereof.

2. A substantially pure NF-AT120 protein or peptide thereof of Claim 1.

3. A protein or peptide of Claim 2, selected from the group consisting of:
   a) a protein from a mammal, including a human;
   b) a protein comprising at least one polypeptide segment of SEQ ID NO: 1 through 5;
   c) a protein which exhibits a post-translational modification pattern distinct from natural NF-AT120 protein; and
   d) a protein which exhibits at least one of the features disclosed in Table 1.

4. A composition comprising a protein of Claim 2, and a pharmaceutically acceptable carrier.

5. An antibody of Claim 1.

6. An antibody of Claim 5, wherein:
   a) said NF-AT120 protein is a mammalian protein, including a human;
   b) said antibody is raised against a peptide sequence of SEQ ID NO: 1 through 5, 35, 37, 39, or 41;
   c) said antibody is a monoclonal antibody; or
   d) said antibody is labeled.

7. A nucleic acid of Claim 1.

8. A nucleic acid of Claim 7, wherein said nucleic acid comprises a sequence of SEQ ID NO: 6 through 24, 34, 36, 38, or 40.

9. A kit comprising:
   a) a substantially pure NF-AT120 protein or fragment;
   b) an antibody which specifically binds an NF-AT120 protein; or
   c) a nucleic acid encoding an NF-AT120 protein or peptide.

10. A method of modulating physiology or development of a cell comprising contacting said cell with a modulator of an NF-AT120 protein.

11. A method of Claim 10, wherein said modulator is an antibody against a mammalian NF-AT120 protein.

12. A method of Claim 10, wherein said cell is a hematopoietic cell, including a lymphoid cell.

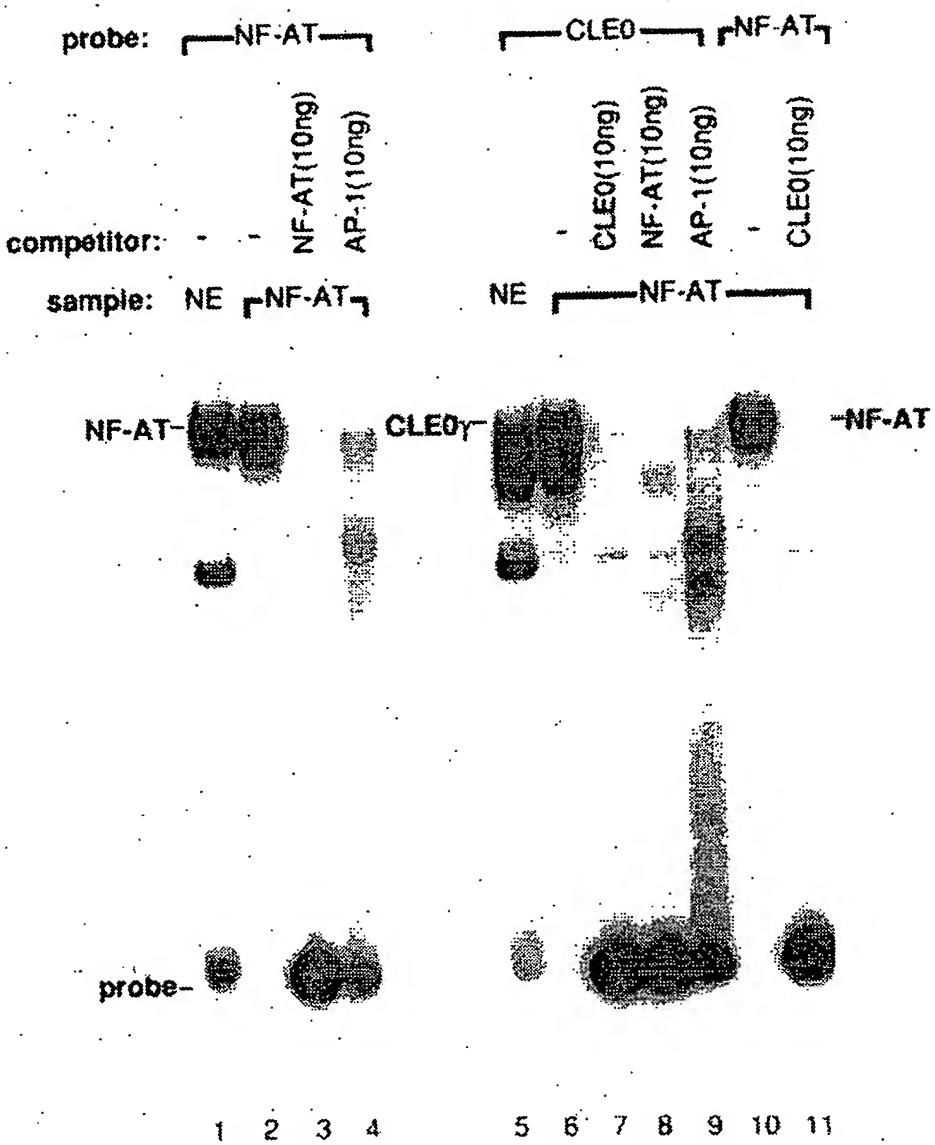
probe:
NF-AT
CLE0
NF-AT
competitor:
-
-
NF-AT(10ng)
AP-1(10ng)
-
CLE0(10ng)
NF-AT(10ng)
AP-1(10ng)
-
CLE0(10ng)
sample:
NE
NF-AT
NE
NF-AT
NF-AT
CLE0γ
NF-AT
probe
1 2 3 4 5 6 7 8 9 10 11

Figure 1

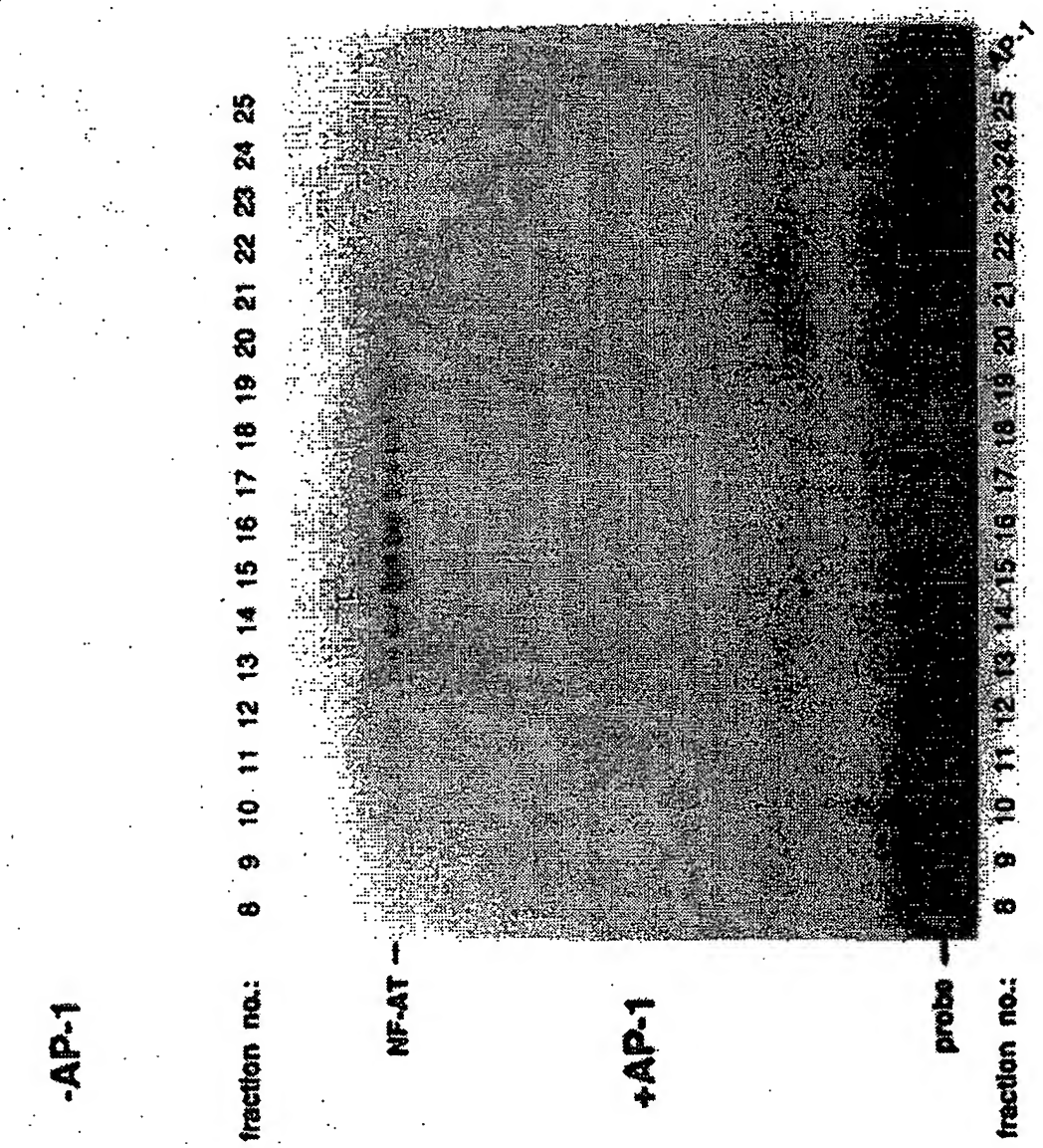

Figure 2A

Figure 2B

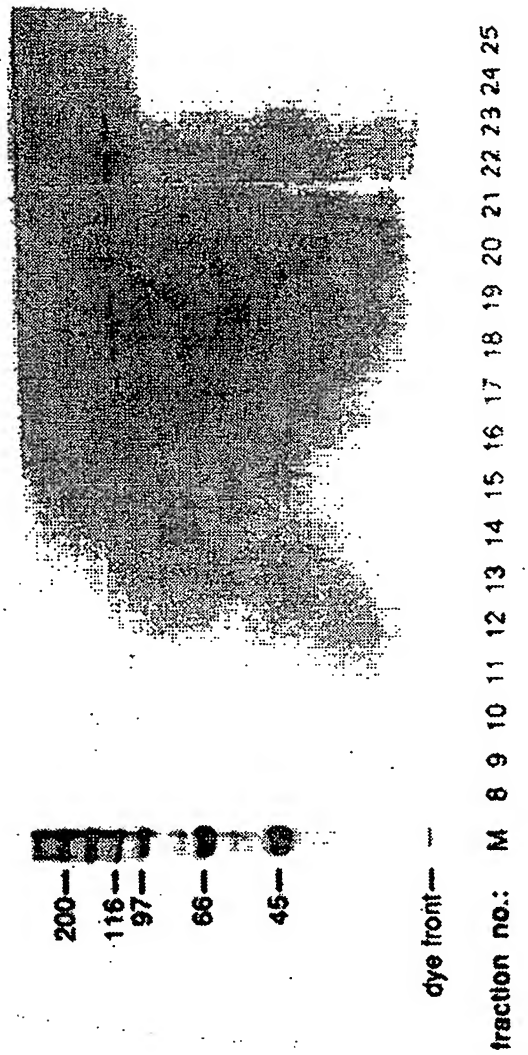
200
116
97
66
45
dye front
fraction no.: M 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25

Figure 2C

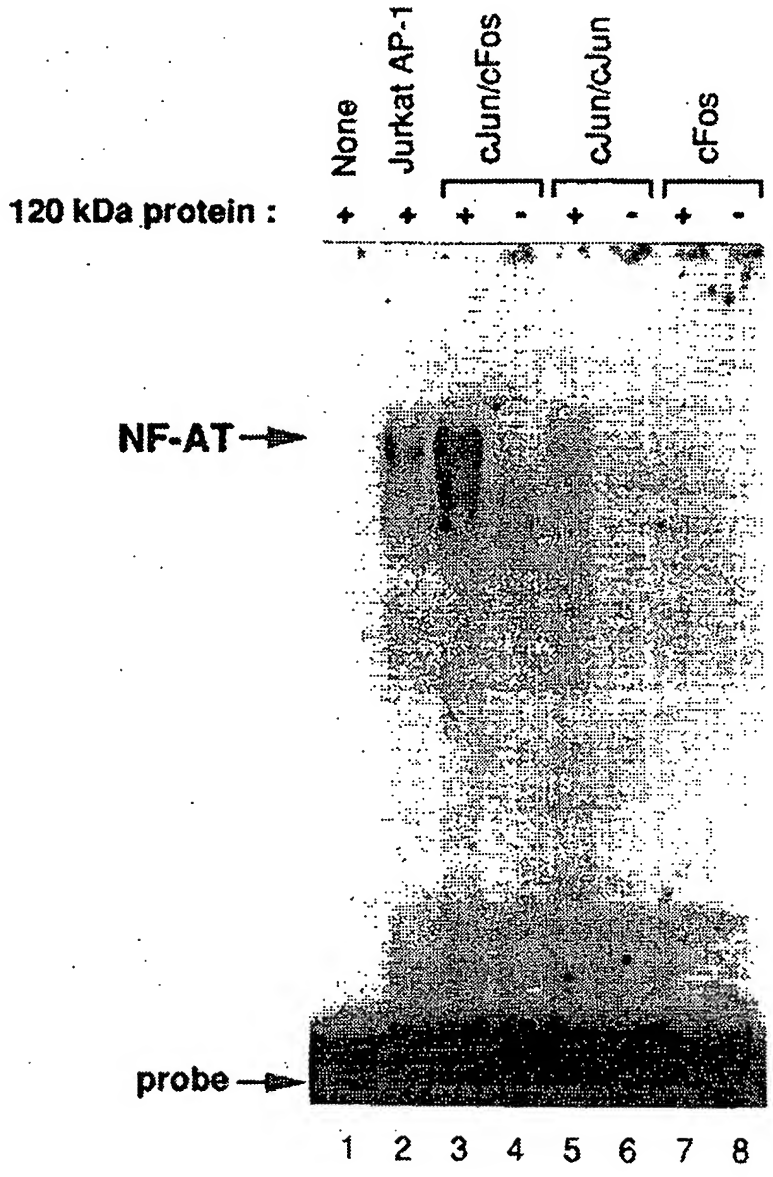
None
Jurkat AP-1
cJun/cFos
cJun/cJun
cFos
120 kDa protein :
+ + + - + - + -
NF-AT
probe
1 2 3 4 5 6 7 8

Figure 3

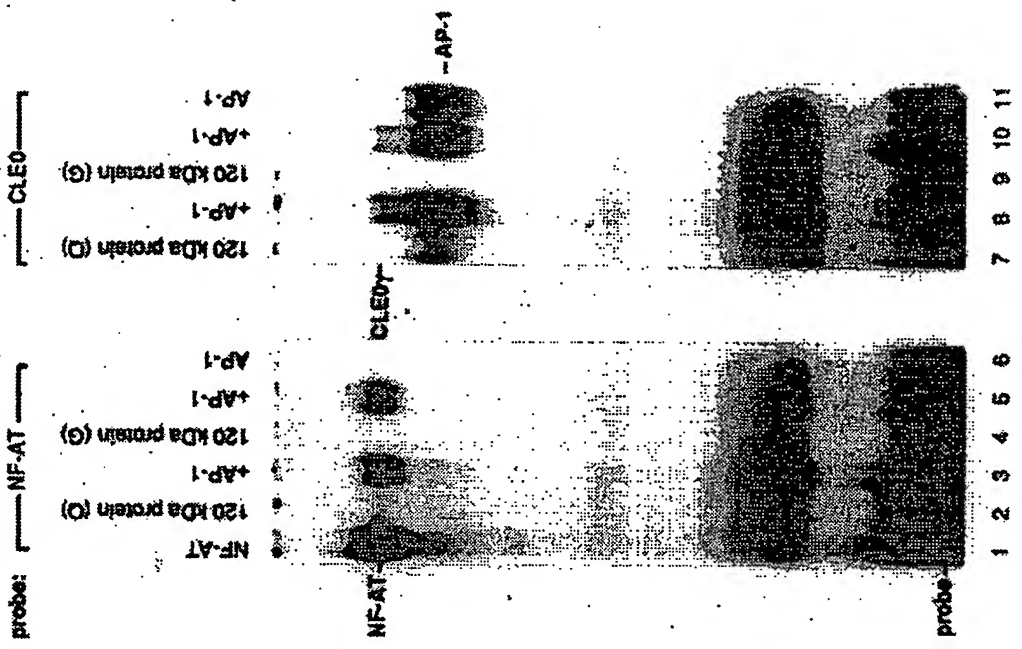

Figure 4B

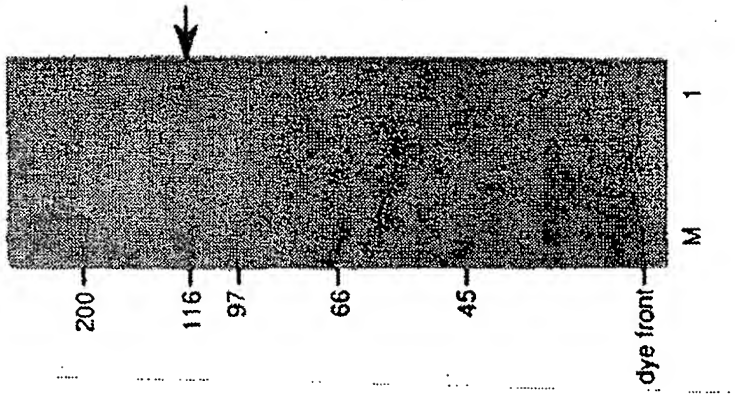

Figure 4A

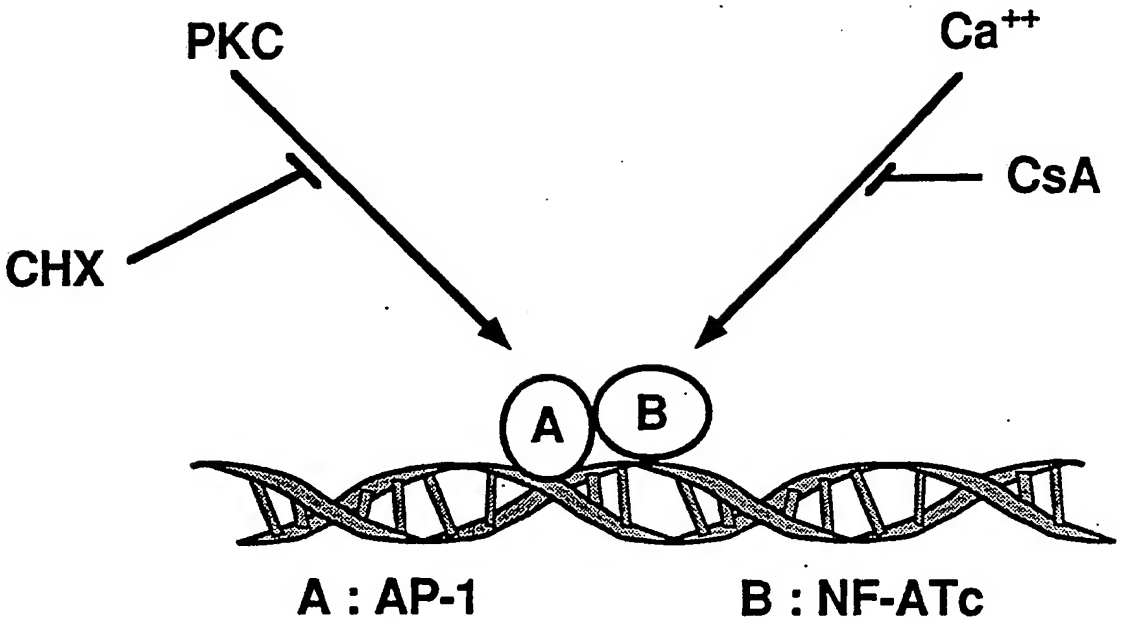

Figure 5

INTERNATIONAL SEARCH REPORT

Interna. .l Application No
PCT/US 94/07297

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/12 C07K16/18 C07K14/47 A61K38/17 G01N33/68
G01N33/53 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JOURNAL OF BIOLOGICAL CHEMISTRY, vol.268, no.4, 5 February 1993, BALTIMORE, MD US pages 2917 - 2923 J.P. NORTHROP ET AL 'Characterization of the Nuclear and cytoplasmic components of the Lymphoid-specific Nuclear Factor of activated T cells (NF-AT) complex' cited in the application see the whole document --- -/--	1-3

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
21 November 1994	02 -12- 1994

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax (+31-70) 340-3016	Le Cornec, N

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS, vol.196, no.2, 29 October 1993, DULUTH, MINNESOTA US pages 737 - 744 H. TOKUMITSU ET AL 'Purification of the 120 kDa component of the human Nuclear Factor of Activated T cells (NF-AT)' see the whole document	1-4
E	WO,A,94 15964 (DANAFARBER CANCER INSTITUTE , INC.) 21 July 1994 * see the whole document especially sequence ID no 12 and the claims *	1-7,9
P,X	SCIENCE, vol.262, 29 October 1993, LANCASTER, PA US pages 750 - 754 P.G. MCCAFFREY ET AL 'Isolation of the cyclosporin-sensitive T cell transcription factor NFATp' see the whole document	1-3,5,6
X	WO,A,93 04203 (THE BOARD OF TRUSTEES OF THE LELAND STANFORD JR. UNIVERSITY) 4 March 1993	1,2,4,7, 9
A	see the whole document	10
A	JOURNAL OF BIOLOGICAL CHEMISTRY, vol.268, no.5, 15 February 1993, BALTIMORE, MD US pages 3747 - 3752 P. MCCAFFREY ET AL 'NF-ATp , a T lymphocyte DNA-binding proteinthat is a target for calcineurin and immunosuppressive drugs' cited in the application see the whole document	1-3
P,X	NATURE, vol.369, 9 June 1994, LONDON GB pages 497 - 502 J.P. NORTHROP ET AL 'NF-AT components define a family of transcription factors targeted in T-cell activationt' * see the whole document especially figure 1 *	1,2,4-8
A	NATURE, vol.356, 30 April 1992, LONDON GB pages 801 - 804 J. JAIN ET AL 'Nuclear factor of activated T cells contains Fos and Jun' cited in the application see the whole document	1,7,9

3

INTERNATIONAL SEARCH REPORT

Information on patent family members

Interna... al Application No
PCT/US 94/07297

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO-A-9415964	21-07-94	NONE	
WO-A-9304203	04-03-93	CA-A- 2114434	04-03-93